

Evaluation

Arnold, Maik; Mayer, Thomas

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Arnold, M., & Mayer, T. (2010). Evaluation. In A. Weidemann, J. Straub, & S. Nothnagel (Hrsg.), *Wie lehrt man interkulturelle Kompetenz? Theorien, Methoden und Praxis in der Hochschulausbildung; ein Handbuch* (S. 525-562). Bielefeld: transcript Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-342092>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Arnold, M. & Mayer, T. (2010). Evaluation. In A. Weidemann, J. Straub, & S. Nothnagel (Hrsg.), *Wie lehrt man interkulturelle Kompetenz? Theorien, Methoden und Praxis in der Hochschulausbildung* (S. 525-562). Bielefeld: transcript.

Bei diesem Dokument handelt es sich um eine Kopie des gedruckten Beitrags im o.g. Sammelband. Das Dokument unterliegt dem Copyright der Verfasser. Eine elektronische Veröffentlichung erfolgt aufgrund schriftlicher Genehmigung durch den Transcript Verlag.

5 Evaluation

MAIK ARNOLD UND THOMAS MAYER

Unter Evaluation soll hier die Bewertung der Wirksamkeit und Effizienz, Qualität und Güte, des Nutzens und der Kosten von Maßnahmen, Programmen, Prozessen oder Produkten verstanden werden. Je nach Anwendungsbereich oder Fachgebiet wird der zu evaluierende Gegenstand eingegrenzt.¹ Hier betrachten wir Studienangebote zur Vermittlung interkulturellen Kompetenz im Hochschulkontext und grenzen damit Evaluation auf die Bewertung von „psychologischen Interventionsmaßnahmen“ (Hager/Patry/Brezing 2000) ein, welche auf eine Verbesserung von Fähigkeiten, Vertiefung von Wissen, Leistungssteigerung der beteiligten Akteure und effektivere Kommunikation und Kooperation mit anderen Personen ausgerichtet sind (Kealey/Protheroe 1996) – und zwar speziell in interkulturellen Konstellationen. In jedem Fall ist das Verständnis von Evaluation maßgeblich vom zugrunde gelegten theoretischen Bezugsrahmen und von der methodischen Vorgehensweise abhängig (vgl. u.a. Arnold 2005; Mendenhall/Stahl/Ehnert/Oddou/Osland/Kühlmann 2004). Die weiteren Ausführungen widmen sich der Frage, wie und in welcher Form in der Hochschullehre anwendbare Verfahren und Methoden zur Ausbildung interkultureller Kompetenz (siehe Teil 4 i.d.B.) im Sinne der Qualitätssicherung (siehe Blüml i.d.B., Kap. 3.3.) evaluiert werden können. Dafür soll eine theoretisch fundierte und methodisch verfahrenende Evaluationsstrategie vorgeschlagen werden.

1 Die Deutsche Gesellschaft für Evaluation (DeGEval) listet auf dem Deckblatt ihrer „Standards für Evaluation“ in loser Folge eine Sammlung verschiedenster Anwendungsbereiche auf, u.a. Hochschulen, Verwaltungen, Schulen, Gesundheitswesen, soziale Dienstleistungen, Aus- und Weiterbildung, betriebliche und berufliche Bildung, Entwicklungszusammenarbeit, Forschungs-, Technologie- und Innovationspolitik, Umweltpolitik, Kommunikation und Medien, Stadt- und Regionalentwicklung; dies spiegelt letztlich die Breite ihrer Mitgliedsinteressen wider, die sich auch in entsprechenden Arbeitskreisen niederschlagen (DeGEval – Deutsche Gesellschaft für Evaluation e.V. 2008).

Nach einer kurzen begrifflichen Einführung zum allgemeinen Verständnis von ‚Evaluation‘ bzw. dem Praxis- und Forschungsfeld ‚Evaluieren‘ im zweiten Abschnitt konzentrieren wir uns im dritten Teil auf die Planung, Durchführung und Implementierung von Evaluationsstrategien in interkulturell ausgerichteten Aus- und Weiterbildungsmaßnahmen an Hochschulen. Dabei stehen die Fragen im Vordergrund: *Wer soll was mit wem, wozu, wann, wie, womit* und unter welchen *Bedingungen* evaluieren? Diese Fragestellungen leiten die weiteren Ausführungen an, sie sind jedoch nicht abschließend bzw. grundsätzlich beantwortbar, sondern bei jeder Evaluation wieder neu zu stellen und zu beantworten. In die Darstellung fließen zudem aktuelle Forschungsbefunde zur Einschätzung der Wirkungen von Interventions- und Qualifikationsmaßnahmen zur Vermittlung interkultureller Kompetenz aus anderen (meist wirtschaftlichen) Kontexten ein (vgl. Landis/Bennett/Bennett 2004). Es ist dabei auch kritisch zu fragen, welche Erfahrungen bisher mit welchen forschungspraktischen Vorgehensweisen in der Untersuchung der Effektivität von (interkulturellen) Qualifizierungsmaßnahmen gemacht wurden, welche Schwierigkeiten und Probleme im interessierenden Bereich erwartbar sind, wie diese überwunden werden können, und welche Forschungsperspektiven und -ergebnisse aus der interkulturellen Trainings- und Evaluationsforschung sich überhaupt auf die Lehr-Lernpraxis an Hochschulen übertragen lassen. In diesem Zusammenhang wird abschließend auf die Grenzen und mögliche Forschungsperspektiven der Evaluation in der Hochschullehre einzugehen sein.

Begriff und Konzept

In der Literatur existiert eine beinahe unüberschaubare Vielfalt an Definitionen des Begriffs ‚Evaluation‘. Begriffsbestimmungen variieren aufgrund der zu beurteilenden Gegenstände, der dabei verfolgten Ziele und Strategien sowie der eingesetzten Methoden und Verfahren erheblich. Von Evaluieren (*lat.* *valor* ‚Wert‘; *valere* ‚wert sein‘; *e-/ex-valere* ‚einen Wert aus etwas ziehen‘) wird i.d.R. immer dann gesprochen, wenn der Wert eines zu beurteilenden Gegenstandes bestimmt bzw. eingeschätzt werden soll (Scriven 1991). Wenn gleich der Begriff in der alltäglichen Kommunikation kaum Verwendung findet, stellt Evaluieren auch eine alltägliche Handlungspraxis dar. Mit der menschlichen Wahrnehmung ist immer auch eine Bewertung verbunden, z.B. immer dann, wenn etwas als gut oder schlecht, schön oder hässlich, erlaubt oder nicht erlaubt, nützlich oder wertlos empfunden wird.

Kommen im Prozess der Festlegung und Beurteilung des Wertes einer Sache explizit (sozial-)wissenschaftliche Forschungsmethoden und -verfahren zum Einsatz, so kann man von ‚Evaluationsforschung‘ sprechen. Eine viel zitierte Definition stammt von Rossi und Freeman (1993: 5): „Evaluation re-

search is the systematic application of social research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs“. In ähnlicher Weise fasst auch die Deutsche Gesellschaft für Evaluation die Bewertung eines Programms als

„die systematische Untersuchung des Nutzens oder Wertes eines Gegenstandes. Solche Evaluationsgegenstände können z.B. Programme, Projekte, Produkte, Maßnahmen, Leistungen, Organisationen, Politik, Technologien oder Forschung sein. Die erzielten Ergebnisse, Schlussfolgerungen oder Empfehlungen müssen nachvollziehbar auf empirisch gewonnenen qualitativen bzw. quantitativen Daten beruhen.“ (DeGEval – Deutsche Gesellschaft für Evaluation e.V. 2008: 13)

Die Evaluationsforschung wird – im Gegensatz zur deskriptiven und explanativen Grundlagenforschung (Daumenlang/Altstötter/Sourisseaux 1999) – als „Anwendungsforschung“ (Patton 2002) oder als „Begleitforschung“ (Kromrey 1988) aufgefasst, weil sie die praktische Umsetzung oder Anwendung von Maßnahmen (begleitend) untersucht und dabei meist an Ziele und Rahmenbedingungen der Auftraggeber gebunden ist und unter Ergebniszwang steht. Bei der Evaluationsforschung handelt es sich um ein disziplinenübergreifendes Feld, das sich mittlerweile u.a. in den Erziehungswissenschaften und Wirtschaftswissenschaften, der Psychologie und Pädagogik fest etabliert hat und zunehmend auch in anderen Forschungsbereichen thematisiert wird.

Eine genaue Begriffsbestimmung wird dadurch erschwert, dass der Begriff ‚Evaluation‘ häufig synonym mit anderen Begriffen wie ‚*program evaluation*‘, ‚Erfolgskontrolle‘, ‚Wirkungsanalyse‘, ‚(Bildungs-)Controlling‘, ‚Effizienzforschung‘, ‚Qualitätskontrolle‘ oder ‚Begleitforschung‘ verwendet wird (Beywl/Schobert 2000). Die Evaluation von (sozialen) Interventionsmaßnahmen wird im Allgemeinen als ‚Programm-Evaluation‘ (Rossi/Freeman 1993) bezeichnet. Bei dem wissenschaftlichen Konzept ‚Evaluation‘ handelt es sich (1) um eine systematische, prozesshafte Vorgehensweise, die dazu dient, (2) unter Einsatz von Methoden und Verfahren der Sozialwissenschaften (3) Informationen über das zu beurteilende Programm oder die Maßnahme zu gewinnen, welche (4) zur Planung, Konzeption und Implementierung sowie für wichtige Entscheidungen über die Veränderung, Fortführung oder Liquidierung eines Programms bzw. einer Maßnahme und (5) zur Bewertung damit verwobener Handlungsalternativen herangezogen werden können (Kardorff 2007; Wottawa/Thierau 2003). Von Evaluation wird im Folgenden immer dann gesprochen, wenn Bewertungen systematisch vorgenommen werden, um damit ein bestimmtes Ziel zu verfolgen.

Wenn man danach fragt, welche intendierten und nicht-intendierten Wirkungen eine Maßnahme oder ein Programm entfaltet oder wie deren konkrete Ergebnisse oder Effektivität beurteilt werden sollen, ist dies von der Werte-

position des Beurteilers abhängig (Beywl 2006). Der Modus des Bewertens kann unterschieden werden von der Beschreibung eines Gegenstandes. Eine Beschreibung enthält keine Beurteilung, Handlungsempfehlung oder Normierung (vgl. Arnold 2005, 2009). Sie beschränkt sich auf die Darstellung dessen *was war, ist oder sein wird*. Bewertungen setzen dagegen Kriterien und Ansprüche voraus, legen einen Maßstab an, definieren und beurteilen Qualitäten.

Evaluation ist immer auch Teil eines „Entwicklungszyklus“ von Lernsystemen (Mayer 2001: 327ff.). Diese Modellvorstellung sieht vor, dass Bewertungsprozesse stets auf allen Ebenen eines Lernsystems vorgenommen werden – von der Makro-Ebene, die das Lernsystem als Ganzes umfasst (z.B. ein kompletter Hochschulstudiengang) bis zur Mikro-Ebene, bestehend aus den Einzelinteraktionen als kleinsten Bestandteilen eines Lernsystems (z.B. Frage eines Dozenten – Antwort eines Studierenden). Auf allen Ebenen, von der Makro- über ggf. mehrere Meso-Ebenen bis zur Mikro-Ebene wiederholt sich ein kontinuierlicher Entwicklungs- und Veränderungszyklus aus Aktivitäten (1) der *Entstehung* (auch Planung, Konzeption, Design, Herstellung), (2) der *Durchführung* (die Lehre selbst) und (3) der *Bewertung*. Die Erkenntnisse aus den Bewertungen können zu *Veränderungen* (entspricht neuerlicher Entstehung) führen, womit der Zyklus sich schließt.

Am Beispiel eines Studiengangs zur Ausbildung interkultureller Kompetenz könnte der Zyklus auf der Makro-Ebene so aussehen: (1) Entstehung: Ein interdisziplinäres Gründungsgremium entwickelt zunächst eine Mantelnote, einen Rahmenstudienplan, den Studienablaufplan und Modulbeschreibungen sowie die Studien- und Prüfungsordnungen, welche an einschlägige Lehrveranstaltungskonzepte anknüpfen und am übergeordneten Ziel des Studienangebots ausgerichtet sind. (2) Durchführung: Das Studienangebot wird mit einer Pilotgruppe von Studierenden erstmalig durchgeführt. (3) Bewertung: Der Studiengang wird im Rahmen eines Akkreditierungsprozesses einer eingehenden Prüfung unterzogen. Das Ergebnis dieser (Über-)Prüfung führt zu Veränderungen in den Studiendokumenten (1), die in eine zweite Durchführung (2) münden. Diese wird dann idealerweise mit einer Begleitforschung evaluiert (3), daraufhin weiter entwickelt (1), und in der dritten Durchführung (2) installiert die Studienleitung ein kontinuierliches Verbesserungssystem. Derartige Entwicklungszyklen finden zugleich – explizit oder implizit – auch auf den tieferen Ebenen des Lernsystems statt: So kann sich eine Dozentin, die mit der Konzeption und Durchführung z.B. eines Seminars beauftragt ist, im Zyklus aus Entstehung, Durchführung, Bewertung und Verbesserung wiederfinden – hinsichtlich ihrer Lehrveranstaltungen als Ganzes, aber auch für jede einzelne Sitzung. Das setzt sich fort, bis schließlich auf der Mikro-Ebene für die angenommen kleinste Einheit der Einzelinteraktion mit dem Lerner unterstellt wird, dass z.B. eine Frage (1) durch Vorüberlegungen der Dozentin zur Formulierung der Frage *entsteht*, (2) in Form der Frage der Dozentin und

der Antwort des Studierenden *durchgeführt* und (3) durch das Urteil der Dozentin über die Geltung, Eignung und Reichweite ihrer Frage *bewertet* wird. Erhält die Dozentin auf ihre Frage keine Antwort, könnte dies dazu führen, dass sie ihre Frage überdenken und sich eine andere einfallen lassen muss, womit der Zyklus von vorn beginnt. Mit diesem Modell soll verdeutlicht werden, dass Bewertungen mehr oder weniger explizit oder implizit, systematisch oder unsystematisch, gezielt oder planlos, objektiv oder subjektiv, bewusst oder unbewusst, summativ-effektorientiert oder formativ-prozessorientiert auf allen operativen Ebenen der Lehr-Lern-Prozesse eines Lernsystems erfolgen können. *Bewertung* als informeller Prozess der Interpretation von Erfahrungen und Handlungen, Erwartungen und Interaktionen in einem Lernsystem durch einen der beteiligten Akteure findet kontinuierlich statt, während *Evaluation* ein explizit planvolles Vorgehen darstellt.

Bewertungen spielen eine wichtige Rolle bei der Qualitätssicherung und der Wirkungseinschätzung von Lehrprozessen sowie der Beurteilung des Lernerfolgs. Einerseits werden die Dozenten bzw. Lehrenden die Wirkungen ihrer gezielten Interaktionen mit den Studierenden mehr oder weniger bewusst einschätzen und beurteilen, um danach das eigene Handeln auszurichten bzw. anzupassen. Andererseits wirken auch die Universitäts-, Fakultäts- und Institutsleitenden darauf hin, dass die Qualität der Lehre gesichert sowie ein geeignetes Qualitätssicherungssystem etabliert wird. Der Vorgang des Bewertens spielt sich im Spannungsverhältnis zweier eng miteinander verwobenen Perspektiven ab: Bewertungen stützen sich einerseits auf die – idealerweise empirisch fundierten und theoretisch gesättigten – Sichtweisen und Einschätzungen, Wertorientierungen und Interpretationen der Evaluierenden und andererseits auf die Aussagen der Befragten der Evaluation. In aller Regel werden im Evaluationsprozess auch ethisch-moralische Bewertungen vorgenommen, wenn sich die Bewertungsvorgänge z.B. auf eine Beeinflussung, Steuerung und Kontrolle menschlicher Handlungen sowie damit verbundener Orientierungen, Einstellungen und Überzeugungen beziehen (Wottawa/Thie-rau 2003).

Nach diesen begriffsgeschichtlichen und theoretischen Vorbemerkungen skizzieren die folgenden Ausführungen eine Vorgehensweise für die Evaluation von Lehr-Lern-Systemen in der Hochschulausbildung, die das Ziel verfolgen, interkulturelle Kompetenz zu vermitteln und zu vertiefen.

Evaluation von Verfahren zur Vermittlung interkultureller Kompetenz in der Hochschule

Aufgrund der oben dargestellten Heterogenität von Begriff und Praxis der Evaluation ergibt sich die Notwendigkeit, im konkreten Einzelfall zu prüfen und zu entscheiden, wie ein Evaluationsvorhaben geplant, konzipiert und ent-

sprechend durchgeführt werden soll, d.h. welche Ziele und Vorgehensweisen gewählt werden sollen, um die Wirkungen ausgewählter Lehrmethoden beurteilen zu können. Die Evaluation von interkulturell ausgerichteten Studienangeboten kann sich auf eine Vielzahl von Ansätzen, Manualen, Kriterien- und Empfehlungskatalogen stützen, um Merkmale guter Lehre und Kriterien für ihre Beurteilung aufzustellen (u.a. Becker-Richter/Habel/Rinke/Tegethoff 2002; Spiel 2001). Bisherige Lehrevaluationen haben neben der Frage nach der Erreichung explizit definierter Lehr- und Lernziele allerdings vorwiegend nach den Rahmenbedingungen der Lehre (z.B. Bewertungen didaktischer Aspekte, Betreuungsverhältnisse und Administrationsaufwand) gefragt. Im Rahmen der Erreichung der Lehr-/Lernziele und der Qualität der Rahmenbedingungen kann zur Bewertung der Qualität der Hochschullehre insbesondere die Frage nach der Vermittlung von Schlüsselqualifikationen gestellt werden. Als Merkmale einer ‚guten‘ Lehre wurden neben Fachkompetenzen auch der Erwerb bzw. die Entwicklung von (berufs-)praxisrelevanten Kompetenzen identifiziert, welche aus den spezifischen Bedürfnissen des Arbeitsmarktes hervorgehen; ebenso wurden auch analytische, methodische, soziale und (selbst-)reflexive Kompetenzen herangezogen. Zunehmend wird in diesem Sinne auch ein Bedarf zur Entwicklung interkultureller Kompetenz in durch kulturelle Differenzen, Andersheit und Fremdheit gekennzeichneten Anwendungs- und Praxisfeldern gesehen. Bei der Planung, Durchführung und Implementierung von Evaluationsmaßnahmen im Kontext von Studium und Lehre am Lernort Hochschule kann darüber hinaus auch auf die Forschungsergebnisse und -perspektiven verschiedener Evaluationsstudien der Lehr-/Lernforschung im Rahmen der Vermittlung interkultureller Kompetenz an Schulen und in der Erwachsenenbildung zurückgegriffen werden, welche z.B. den Erwerb bzw. die Entwicklung interkultureller Kompetenz im Rahmen der schulischen Sozialisation zum Thema ihrer Untersuchungen machen (z.B. Bertels/Eylert/Lütke/de Vries 2004). Diese Studien bedürfen jedoch sowohl hinsichtlich ihrer konzeptuellen Vorgehensweise als auch in Bezug auf die systematische Entwicklung von Evaluationsmaßnahmen und den Einsatz von kombinierten Evaluationsinstrumenten noch entsprechender Weiterentwicklungen.

Aus dem zentralen Anliegen dieses Handbuchs leitet sich die für die folgenden Überlegungen wesentliche Frage ab, wie gängige Verfahren und Methoden der Vermittlung interkultureller Kompetenz in der Hochschulausbildung (siehe Teil 4 i.d.B.) beurteilt und optimiert werden können. Die Entwicklung eines problemadäquaten, die organisationalen und institutionellen Bedingungen des Lernumfeldes berücksichtigenden Evaluationskonzeptes orientiert sich an den folgenden acht (Teil-)Fragekomplexen, die geeignet erscheinen, die Evaluation von Maßnahmen und Programmen im Allgemeinen und speziell von interkulturell ausgerichteten Studien- und Weiterbildungsan-

geboten zu kennzeichnen. Mitunter sind auch allgemeine Bemerkungen zur Gestaltung von Evaluationsmaßnahmen notwendig, da ein grundlegendes Verständnis von dem, was Evaluation in der Hochschullehre eigentlich leisten kann, erst erarbeitet und kritisch diskutiert werden muss, bevor man überhaupt an eine Übertragung von Erkenntnissen aus anderen Untersuchungs- und Anwendungsbereichen denken kann.

1. *Was soll evaluiert werden? – Evaluationsgegenstand:* Am Anfang steht die Frage, was, überhaupt evaluiert werden soll, also welches Programm, Projekt oder welche Intervention Gegenstand der Bewertung sein soll. Im Folgenden wird der Gegenstand eingegrenzt auf soziale Interventionsmaßnahmen der *Hochschullehre*, die aufgrund ihrer inhaltlichen Ausrichtung auf die *Vermittlung interkultureller Kompetenz* abzielen.
2. *Wozu soll eine Evaluation durchgeführt werden? – Evaluationsziele:* Die Planung, Durchführung und Implementierung einer Evaluation ist von ihren konkreten Zielen abhängig. Darum wird hier mit einer Zieltaxonomie (s.u.) ein systematisches Raster vorgelegt, um zu transparenten Entscheidungen über Evaluationsziele für Lernsysteme gelangen zu können.
3. *Wie soll die Evaluation durchgeführt werden? – Evaluationsdesign:* Hier werden Fragen geklärt, wie die Evaluation formalstrukturell ausgestaltet werden muss, und wie konzeptuelle Rahmenbedingungen zur Einschätzung der Güte, Qualität und Validität der Evaluationsergebnisse geschaffen werden können.
4. *Womit wird evaluiert? – Erhebungs- und Auswertungsinstrumente:* Auf Grundlage der Entscheidung des grundlegenden Evaluationsdesigns geht es um die einzusetzenden Methoden und Instrumente der Datenerhebung, -aufbereitung und -auswertung.
5. *Wann wird evaluiert? – Evaluationszeitpunkt:* Auf Grundlage des Designs, der Evaluationsziele und der verwendeten Methoden werden die Zeitpunkte für den Untersuchungsablauf festgelegt.
6. *Wer soll die Evaluation durchführen? – Evaluator/in bzw. Evaluatorenteams:* Es muss geklärt werden, welche Personen die Evaluation durchführen, und es muss entschieden werden, in welcher Form sich diese an der Evaluation beteiligen bzw. welche Rolle sie im Evaluationsprozess einnehmen sollen.
7. *Wen betrifft die Evaluation? – Akteure, Teilnehmer und Stakeholder:* Es müssen geeignete Forschungspartner gefunden werden, die bereit sind, an der Untersuchung teilzunehmen. Je nach Fragestellung kommen sowohl die Teilnehmer der Interventionsmaßnahme als auch andere Interessensgruppen bzw. Akteure im Kontext Hochschule (siehe Weidemann/Nothnagel i.d.B., Kap. 3.1) in Betracht.
8. *Unter welchen Bedingungen wird evaluiert? – Evaluationsstandards und Gütekriterien:* Bei der Planung, Durchführung und Implementierung von

Evaluationsstrategien müssen Qualitätsstandards und Gütekriterien beachtet werden, die unmittelbaren Einfluss auf die Güte und Aussagekraft der Evaluationsergebnisse haben.

Evaluationsgegenstand

Gegenstand der Evaluation können verschiedene Interventionsmaßnahmen im Bildungsbereich sein. Für den hier betrachteten Kontext der Hochschullehre kann das Evaluationsobjekt mithilfe von vier Komponenten inhaltlich bestimmt werden: (1) Die Evaluation findet im Lehr-Lernsystem und *Kontext* ‚Hochschule‘ statt. (2) Der konkret zu evaluierende Gegenstand stellt eine *Maßnahme der Hochschullehre* dar, in der (3) bestimmte *Methoden und Verfahren* mit dem Ziel der (4) *Vermittlung interkultureller Kompetenz* eingesetzt werden. Diese vier gegenstandsverankerten Merkmale der Evaluation bedürfen einer näheren Bestimmung.

Hochschule als Lernsystem und Kontext der Evaluation

Die Hochschule stellt den Ort und den spezifischen Lehr-Lernkontext der Evaluation dar. Für das komplexe Gebilde ‚Hochschule‘ gibt es weder ein präzise definierbares ‚Programm‘ mit einheitlich vorgegebenen Zielen, damit zusammenhängenden konkreten Maßnahmen und bestimmten Kriterien zur Erreichung dieser Ziele, noch existiert ein universell abgrenzbares ‚Produkt‘, dessen Güte, Wert und Qualität mithilfe von standardisierten Methoden und Verfahren der Sozialwissenschaften bestimmt werden könnte (Kromrey 2000). Grundsätzlich stellt eine Hochschule ein Lernsystem dar, in dem spezifische Gruppen mit je unterschiedlichen Zielen und Vorstellungen beschäftigt sind. Alle Beteiligten besitzen – ungeachtet aller Unterschiede – ein tradiertes Handlungs- und Erfahrungswissen darüber, wie diese besondere Organisationsform und deren Infrastruktur aufgebaut ist und weiterentwickelt werden könnte. Die beiden vorrangigen Ziele der Hochschulen sind: erstens Forschung zu betreiben, um damit neues Wissen zu generieren, und zweitens im Rahmen wissenschaftlicher Lehre spezifische Kenntnisse und Fähigkeiten an Studierende zu vermitteln und dafür formale Qualifikationszertifikate, z.B. akademische Grade zu verleihen. Im System Hochschule treffen unterschiedliche Gruppen mit teilweise differierenden Interessen aufeinander: die Studierenden (Lernende), Dozenten (Lehrende), die Universitäts-, Fakultäts- und Institutsleitung sowie verschiedene Stellen der Hochschulverwaltung (siehe Weidemann/Nothnagel i.d.B., Kap. 3.1). Diese Akteursgruppen sind in verschiedenen Organisationsformen, wie z.B. durch Administrativ- und Selbstverwaltungsorgane der Hochschule (z.B. Rektorat bzw. Präsidium, Senat, Studierendenschaft, Studien-, Prüfungsausschüsse und Studiengänge), zu-

sammengeschlossen. Die unterschiedlichen Interessen und Erwartungen bedingen sich gegenseitig und müssen gemeinsam betrachtet werden. Auch wenn die Hochschule hier als Kontext für den eigentlichen Evaluationsgegenstand von Maßnahmen der Hochschullehre betrachtet wird, so ist zu erwähnen, dass auch die Qualität einer Hochschule als gesamtes Lernsystem Gegenstand von Evaluation sein kann.

Maßnahmen der Hochschullehre als konkreter Gegenstand der Evaluation

Der deutsche Wissenschaftsrat (2008) hat in seinen „Empfehlungen zur Qualitätsverbesserung in Studium und Lehre“ konstatiert, dass die Hochschullehre darauf ausgerichtet sein muss, die Studierenden in ihrem Lernen in geeigneter Weise zu unterstützen, wobei spezifische Studienangebote konzipiert werden sollen, die möglichst genau den Fähigkeiten und persönlichen Interessen der Studierenden zu entsprechen haben. Im Vordergrund der Hochschulausbildung steht dabei die Förderung von selbstorganisiertem Lernen durch die Lehrenden und damit auch die Vermittlung von nicht nur fachlichen, sondern auch sozialen und fachübergreifenden Kompetenzen. Die Studierenden sollen beim wissenschaftlichen Lernen und Forschen unterstützt werden sowie sich durch ihr selbstständiges und eigenverantwortliches Handeln weiterentwickeln können. Um dieses Ziel zu erreichen, sind alle Personen im Berufs- und Tätigkeitsfeld Hochschule dazu aufgefordert, geeignete Lehr- und Lernsituationen zu schaffen und Lernprozesse anzuleiten.

Es wird außerdem empfohlen, konkrete Maßnahmen (z.B. Lehrmethoden) einzusetzen, die eine Erfassung und Bewertung des Bildungserfolges an Hochschulen ermöglichen und die Einrichtung eines Informationssystems für Lehre und Studium fördern (ebd.: 77ff.). Eine systematische Qualitätssicherung durch eine Orientierung an ‚Learning Outcomes‘ (siehe Blüml i.d.B., Kap. 3.3) wird in verschiedenen Bereichen der Hochschule gefordert, z.B. bei der Entwicklung, Akkreditierung und Umwandlung von Studiengängen und der Überprüfung der Eignung von Lehr-Lernmethoden für spezifische Lernkontexte. Studiengänge stellen in diesem Zusammenhang formale Bezugsrahmen dar, welche durch spezifische Studien- und Prüfungsordnungen geregelt sind und unter Einsatz von geeigneten Lehr- und Lernformen auf Basis einer spezifischen disziplinären Verankerung umgesetzt werden (Kromrey 2000; Weidemann/Weidemann/Straub 2007).

Lehr- und Vermittlungsmethoden

Die Evaluation in der Hochschullehre ist eng an die Untersuchung der Wirksamkeit der eingesetzten Lehr- und Vermittlungsmethoden zur Erreichung

von intendierten Lernzielen geknüpft. In der Hochschulausbildung kommen üblicherweise die klassischen Lehrformen wie Vorlesung, Seminar, Tutorium, Übung und das Selbststudium zum Einsatz. In interkulturell ausgerichteten Studienangeboten (zu deren Funktionen und Organisationsformen siehe Otten i.d.B., Kap. 3.2) werden zudem aus dem breiten Spektrum an Vermittlungsmethoden vermehrt auch Lehrformen wie Simulationen, Planspiele (siehe Strohschneider i.d.B., Kap. 4.2), Feldforschungen bzw. Exkursionen sowie methodisch angeleitete Lehrforschungsprojekte (siehe Weidemann i.d.B., Kap. 4.16) angewendet (für einen Überblick siehe Teil 4 i.d.B.), deren Wirkweisen nur unzureichend geklärt sind. Darüber hinaus kommen außerdem einzelne Lehrbausteine, wie z.B. Videosequenzen (Gieselmann i.d.B., Kap. 4.9; Bauer i.d.B., Kap. 4.10; Oberlik i.d.B., Kap. 4.11), Unterrichtsmaterial (McDaniel/McDaniel/McDaniel 1988) und Auslandsaufenthalte zum Zwecke von Praktika oder Studienvorhaben (Blake/Heslin 1983; Nothnagel i.d.B., Kap. 4.14) zum Einsatz. Die Wahl der geeigneten Methode bzw. anzuwendenden Vermittlungsverfahren variiert je nach Bildungsangebot und -bedarf, ist i.d.R. aber von den Universitätsleitungen bzw. den Akkreditierungsinstitutionen vorgegeben (Otten i.d.B., Kap. 3.2; Weidemann et al. 2007).

Die Entscheidung über den Einsatz einer bestimmten Lehrform bzw. Vermittlungsmethode ist auch abhängig davon, inwieweit die Studierenden für mögliche Berufs- und Tätigkeitsfelder qualifiziert und auf weiterführende, z.T. forschungsorientierte Studiengänge vorbereitet werden sollen. Teilweise lassen sich hier auch Erkenntnisse der Evaluationsforschung insbesondere von interkulturellen Qualifizierungsmaßnahmen heranziehen. Um den Lernerfolg in der Hochschullehre zu sichern, lassen sich verschiedene Vermittlungsansätze miteinander kombinieren, wie z.B. didaktische, kognitive und informatorische mit erfahrungsorientierten, interaktiven und experimentellen Lehrmethoden, wobei letztere als effektiver gelten, weil sie die Studierenden stärker involvieren (z.B. Bolten 2000). Dennoch entfalten „interaktive Übungen (wie Rollenspiele) [...] ohne ausreichendes kulturelles Wissen und Verständnis“ kaum positive Wirkungen (Ehnert 2007: 445; vgl. auch Gannon/Poon 1997).

Vermittlung interkultureller Kompetenz als Lernziel

„Interkulturelle Kompetenz“ als allgemeine „Schlüsselqualifikation“, überfachliche soziale Fähigkeit und Fertigkeit, wird neben dem für verschiedene Berufs- und Tätigkeitsfelder benötigten Fachwissen als eine wesentliche Voraussetzung für effektives und erfolgreiches professionelles Handeln in kulturell heterogenen Studien- und Arbeitsfeldern angesehen (siehe Straub/Nothnagel/Weidemann i.d.B., Kap. 1; Straub 2007). Demzufolge wird diese Kompetenz mittlerweile in verschiedenen interkulturell ausgerichteten Studienangeboten als ein wesentlicher Bestandteil vermittelt, z.T. stellt sie außerdem ein explizier-

tes Studienziel bzw. den zentralen Ausbildungsgegenstand dar (siehe Otten i.d.B., Kap. 3.2; vgl. Straub/Nothnagel 2007; Bergemann/Bergemann 2003). Konkrete qualitative Inhalte und formale Aspekte interkultureller Kompetenz – wie diese in verschiedenen Definitionen und „Komponenten- und Konstituentenmodellen“ (Straub 2007: 41ff.) verankert sind – werden im jeweiligen Kontext i.d.R. durch Rekurs auf andere Theorien, Konzepte und Begriffe identifiziert und spezifiziert. Dieser Umstand bringt meist mit sich, dass einerseits alle aus diesem Konstrukt abgeleiteten und ableitbaren Lernziele äußerst heterogen, vielschichtig und nicht vollständig operationalisierbar sind und dass andererseits Bedeutungs- und Handlungszusammenhänge durch verschiedene disziplinäre Forschungsperspektiven geprägt sind (ebd.).

Die Komplexität des Konstrukts ‚interkulturelle Kompetenz‘, welches in ein interdisziplinär geprägtes Forschungs- und heterogenes Handlungsfeld eingebettet ist, erschwert nicht zuletzt auch die Operationalisierung von Zielerreichungskriterien und die Wahl geeigneter Testverfahren und Instrumente zur diagnostischen Erfassung dieser Kompetenz (z.B. Deller/Albrecht 2007). Stets müssen aus einem umfangreichen Fundus an wissenschaftlichen Tests – die sich aufgrund z.T. nicht explizierter theoretischer Grundlagen und ihrer Validität und Reliabilität erheblich unterscheiden – geeignete Indikatoren ausgewählt werden, um konkrete ‚Learning Outcomes‘ für zu vermittelnde bzw. für zu erweiternde und zu vertiefende (interkulturelle) Kompetenzen der Lernenden zu definieren (vgl. Blüml i.d.B., Kap. 3.3). Dabei gilt: Nicht alle Indikatoren können bei einer Evaluation berücksichtigt werden, nicht alle sind zielführend. Einheitliche Lernziele lassen sich weder über das Lernsystem Hochschule hinweg noch in seinen kleinsten Einheiten, wie z.B. in unmittelbaren Dozenten-Studierenden-Interaktionen, aufstellen. Die Beurteilung der Effekte bzw. Wirkungen von Lehrveranstaltungen variiert letztlich in Abhängigkeit von den in die Evaluation einbezogenen abhängigen Variablen, Indikatoren und Prädiktoren (z.B. Wissen, Verhalten, Anpassung, Zufriedenheit, Einstellungen; vgl. Mendenhall et al. 2004). Letztlich müssen ebenso organisationale und institutionelle Bedingungen des Lernumfeldes mit betrachtet werden (siehe Otten i.d.B., Kap. 3.2). Auf die Ableitung von geeigneten und bisher erfassten Zielerreichungskriterien im Rahmen von Evaluationen in der Hochschullehre mit dem Ziel einer Vermittlung interkultureller Kompetenz wird im folgenden Abschnitt eingegangen.

Evaluationsziele

Durch die Eingrenzung auf Evaluationen von Bildungsmaßnahmen der Hochschullehre ergibt sich als häufigste Zielsetzung die Ermittlung ihrer Wirksamkeit bezogen auf ihre intendierten Lehr- oder Veränderungsziele. Jedoch sind die möglichen Zielsetzungen einer Evaluation weiter zu differenzieren. Aus

den Unterscheidungen von Zielen, Funktionen und Zwecken von oder Erwartungen an Evaluationen (Arnold 1992; Bortz/Döring 2006; Jenzer 1992; Will/Winteler/Krapp 1987; Wottawa/Thierau 2003 – zusammenfassend siehe Mayer 2001; Ehnert 2007), kann eine Taxonomie entlang zweier Dimensionen entwickelt werden: Erstens nach dem *Prozess*, den ein Lernsystem durchläuft, und zweitens nach der *Perspektive* und den Interessen der beteiligten Akteure bzw. der Befragten.

Prozess-Dimension

Wie beschrieben durchlaufen Lernsysteme und deren einzelne Komponenten einen ‚Entwicklungszyklus‘ im Sinne einer sukzessiven Optimierung. Dieser Zyklus kann kontinuierlich und auf unterschiedlichen Ebenen der Teile des Lernsystems stattfinden. Nach diesem Prozess-Modell können folgende Ziele unterschieden werden:

- **Entstehungs-Ziele:** Dabei werden Fragen im Zusammenhang mit der Herstellung, Konzeption, Planung und dem Design einer Maßnahme oder ihrer Teile gestellt, z.B. Bedarfsanalyse.
- **Durchführungs-Ziele:** Dabei geht es um Fragen zur Umsetzung der Maßnahme oder ihrer Teile, z.B. Steuerung des Ablaufs.
- **Bewertungs-Ziele:** Dabei geht es um die im engeren Sinne evaluativen, weil bewertenden Fragen an die Maßnahme oder ihre Teile, z.B. Untersuchung der intendierten Effekte.

Perspektiven-Dimension

Neben der Differenzierung prozessbezogener Ziele muss unterschieden werden, wessen Ziele mit einer Evaluation verfolgt werden. Wesentlich ist dies, da sich durchaus Zielkonflikte zwischen den Perspektiven verschiedener Akteure und Teilnehmer ergeben können. Hierbei lassen sich drei Perspektiven unterscheiden:

- **Maßnahmenbezogene Ziele:** Evaluationsziele, die sich auf die zu untersuchende Maßnahme selbst beziehen, z.B. effiziente Gestaltung eines Hochschulstudienangebots.
- **Kontextbezogene Ziele:** An eine Evaluation können zugleich Ziele und Erwartungen geknüpft werden, die sich nicht nur auf die konkrete Maßnahme und ihre Wirkung richten, sondern auf die Erfüllung von übergeordneten Interessen verschiedener Akteure und Stakeholder (s.u.) im Kontext der Maßnahme. So könnte etwa der Träger eines Programms entscheiden wollen, ob seine Investitionen weiterhin lohnend sind, inhaltliche und/oder wissenschaftliche Erkenntnisinteressen befriedigt werden,

oder ob die Maßnahme mit übergeordneten politischen Zielen kompatibel ist.

- **Subjektbezogene Ziele:** Schließlich können mit einer Evaluation Ziele verfolgt werden, welche der Erfüllung von individuellen Interessen der einzelnen Lerner als Teilnehmer der Maßnahme dienen, z.B. benötigt ein Studierender Feedback, um seinen Lernprozess individuell zu steuern.

Tabelle 1: Zieltaxonomie anhand der Dimensionen Prozess (Entstehung, Durchführung oder Bewertung) und Perspektive (maßnahmen-, kontext- oder subjektbezogen)

Prozess		Perspektiven		
		maßnahmenbezogene Ziele	kontextbezogene Ziele	subjektbezogene Ziele
	Entstehungsziele	<ul style="list-style-type: none"> • Bedarfsanalyse für Zielgruppe • Optimierung der Maßnahme • Auswertung von Metaanalysen • pädagogisch-psychologisches Vorgehen 	<ul style="list-style-type: none"> • Anforderungsanalyse • Aufwandkalkulation • Akkreditierung 	<ul style="list-style-type: none"> • individuelle Bedarfsabfrage • Anpassung von Lernzielen • Auswahl von Lerninhalten • Lernerfreundliche Gestaltung von Bedingungen
	Durchführungsziele	<ul style="list-style-type: none"> • Steuerung von Ablauf und Organisation des Lehrprozesses • Ausgestaltung von Lehr-Arrangements • Dokumentation 	<ul style="list-style-type: none"> • fachdidaktische Erkenntnis • Dialog und Transparenz zwischen Akteuren 	<ul style="list-style-type: none"> • Steuerung des individuellen Lernprozesses • Ausrichtung der Motivation • Steuerung der Handlungsorientierung für Lerntransfer
	Bewertungsziele	<ul style="list-style-type: none"> • Zufriedenheit • Lerneffekt • Transfereffekt • situative Veränderungen • Effizienz der Zielerreichung, (ökonomische) Aufwand-Nutzen-Relation • Nachhaltigkeit der Wirkung • Nebenwirkungen • Evaluation der Evaluation und Ergebnisverwendung 	<ul style="list-style-type: none"> • Kontrolle erwünschter Effekte • Legitimation des Aufwands • Legitimation der inhaltlichen Ausrichtung • Entscheidung über die Zukunft der Maßnahme • Meinungsbildung zur Maßnahme • wissenschaftliche Erkenntnis • eigene Ziele von Evaluatoren und/oder deren Auftraggebern • Meta-Evaluation, Meta-Analysen und Evaluationsreviews 	<ul style="list-style-type: none"> • Selbsterkenntnis über aktuelle Kompetenzen • Nachweis formaler Qualifikation (Zertifizierung) • Berechtigung/Zugang zu anschließenden Lehrangeboten

Aus der Kombination dieser Zieldimensionen ergeben sich neun Zielkategorien (siehe Tabelle 1), die im Folgenden konkretisiert werden sollen.

Maßnahmenbezogene Ziele

Die meisten explizit genannten Evaluationsziele beziehen sich auf die zu evaluierende Maßnahme selbst. Daher sollen die maßnahmenbezogenen Ziele zuerst dargestellt werden (um danach die Perspektive zunächst auf den übergeordneten Kontext, in dem die Maßnahme steht, zu erweitern und sie dann schließlich auf die Perspektive des individuellen Teilnehmers zu verengen).

- *Entstehungsziele:* Ein erster wesentlicher Fokus bei der Entstehung eines Lernsystems ist der Bedarf der Zielgruppe. So selbstverständlich eine *Bedarfsanalyse* im Rahmen der Entwicklung eines Lernsystems ist (bzw. sein sollte), so selten wird sie der Evaluation zugerechnet. Zielgruppen für Angebote zur interkulturellen Kommunikation sind häufig heterogen, was eine besondere Herausforderung für die Analyse ihres Bedarfs darstellt (Leenen 2007); dies gilt auch dann, wenn es sich um vermeintlich homogene, rein studentische Zielgruppen eines interkulturell ausgerichteten Studienganges oder Seminars handelt (Weidemann/Nothnagel i.d.B., Kap. 3.1). Nach erfolgter Durchführung können Erkenntnisse aus der Evaluation zur Veränderung und Verbesserung im Sinne einer *Optimierung* der Maßnahme (also neuerliche Entstehung) und/oder zur Entscheidung über deren Weiterführung bzw. Liquidierung verwendet werden (z.B. Kromrey 2000; Stockmann 2006). So kann z.B. die Wahl der Methoden – besonders relevant im Hinblick auf Verhaltens- und Einstellungsänderungen im Kontext interkultureller Kommunikation (Ehnert 2007) – durch das Evaluationsergebnis beeinflusst werden. Die Erstellung oder Auswertung von *Metaanalysen* im Vorfeld einer Neukonzeption liefert systematische Hinweise zur Gestaltung. Bei der Entwicklung und Optimierung von Lehrformen am Lehr- und Lernort Hochschule spielt auch die Gestaltung des *pädagogisch-psychologischen Vorgehens* in den Prozessen und Handlungen der Akteure der Interventionsmaßnahmen eine bedeutende Rolle (z.B. Ergebnis-, Prozess- und Handlungsorientierungen; vgl. Thierau-Brunner/Stangel-Meseke/Wottawa 1999).
- *Durchführungsziele:* Wesentliches Ziel in Bezug auf die Durchführung der Maßnahme ist es, die *Steuerung des Ablaufs* und die *Organisation des Lehrprozesses* zu unterstützen. Insbesondere bei flexiblen Programmen, die eine heterogene Zielgruppe ansprechen, gilt es, den Ablauf zu überwachen (Monitoring), um ihn günstig beeinflussen zu können. Ein weiteres Ziel kann sich im Zuge der Durchführung auf Entscheidungen über die konkrete *Ausgestaltung der Lernarrangements* beziehen. Auch die *Dokumentation* der Durchführung zum Zweck späterer Recherchen oder Ana-

lysen kann zwar ein eigenes Ziel darstellen. Es ist jedoch zu fragen, ob die Dokumentation tatsächlich Selbstzweck oder nur Prozessschritt zur Datensammlung für weitergehende Bewertungsziele (s.u.) ist (Stockmann 2006). Je innovativer eine Maßnahme ist, desto relevanter wird eine umfassende Dokumentation, um womöglich weitergehende Evaluationsziele, die sich ggf. später erst ergeben, verfolgen zu können.

- *Bewertungsziele:* Im Sinne des hier zugrunde gelegten Evaluationskonzeptes ist dies die am häufigsten genannte Zielkategorie. Scriven (1991) hat dafür den Begriff der *summativen* Evaluation geprägt und damit die Untersuchung von Ergebnissen oder Wirkungen nach oder am Ende eines Programms von einer *formativen* Evaluation unterschieden, deren Ergebnisse sich im Laufe der Entwicklung eines Programms an die einzelnen Akteure des Programms richten. In Scrivens Ansatz wird also die Maßnahme an sich bzw. ihre Effekte im weitesten Sinn bewertet. Kirkpatrick (1979; kritisch z.B. Becker 2005) unterscheidet in seinem Evaluationsmodell dabei folgende vier Ebenen: (1) die *Zufriedenheit* oder Akzeptanz der Lernenden (aber auch von anderen Akteuren wie den Dozenten) als bedeutsames, wenn auch schwer objektivierbarer, wenig reliabler und begrenzt valider Indikator für die Qualität einer Maßnahme (Ebene *reaction*); (2) der *Lerneffekt* bei den Lernenden, messbar als unmittelbar nach Durchführung eintretende Wirkung der Maßnahme (Ebene *learning*); (3) der *Transfereffekt* als nachhaltige Wirkung in Anwendung und Umsetzung des Gelernten durch den Lernenden in dessen Alltagspraxis (Ebene *behavior*); (4) die *situative Veränderung* von äußeren Bedingungen durch die Maßnahme als ebenfalls nachhaltiger Effekt (Ebene *results*). Wenn neben der *Effektivität* (also der Relation zwischen erreichtem und angestrebtem Ziel als Zielerreichungsgrad) auch noch die *Effizienz* der Maßnahme (also der Relation zwischen Zielerreichung und Aufwand bzw. Kosten und Nutzen) gemessen werden soll, so wird von Bildungscontrolling und *„return-on-investment“* gesprochen. Dabei kann der investierte Aufwand monetäre oder andere materielle wie auch zeitliche Ressourcen betreffen. Ein zu Grunde liegendes ökonomisches Prinzip kommt z.B. im Rahmen von Personalentwicklungs- und Controllingkontexten (u.a. als Kennzahlenauswertung oder Investitionsrechnung; vgl. Thierau-Brunner et al. 1999) zum Tragen. Die *Nachhaltigkeit der Wirkungen* zu überprüfen, zielt darauf ab, wie *„stabil“* die erreichten Effekte sind. Schließlich stellt sich die Frage, ob – wie z.B. bei der „zielunabhängigen Evaluation“ (Scriven 1991; Kromrey 1988) – auch *Nebenwirkungen*, d.h. nicht-intendierte Effekte, erfasst werden sollen. Bei der Definition von geeigneten und erfassbaren Indikatoren zur Einschätzung der Wirksamkeit, des Lern- und Transfererfolgs im Rahmen der Hochschullehre kann Bezug genommen werden auf die verschiedenen theoretischen Ansätze zur For-

mulierung von Dimensionen, Konstituenten und Komponenten interkultureller Kompetenz, wie z.B. affektive, kognitive und verhaltensbezogene Merkmale (Bolten 2000: 68) oder personen- und kommunikationsbezogene, psychologische und kulturspezifisch Anpassungskriterien (Chen/Sta-rosta 1996; für einen Überblick siehe Rathje 2006; Straub 2007). Eine besondere Art von Bewertungszielen stellt die *Evaluation der Evaluation* selbst dar, d.h. wenn die Qualität der Bewertungsprozesse selbst untersucht wird. Schließlich beeinflusst auch die *Ergebnisverwendung* den Bewertungsprozess: U.a. wird vor und während der Sammlung von Daten und zu Beginn ihrer Auswertung und Interpretation festgelegt, welche Evaluationsergebnisse tatsächlich genutzt werden sollen.

Kontextbezogene Ziele

Die Evaluationsziele, die sich auf den übergeordneten Kontext einer Maßnahme und die entsprechenden Interessen verschiedener Akteure beziehen, werden gelegentlich auch als tiefer liegender ‚Zweck‘ einer Evaluation bezeichnet. Der Kontext einer einzelnen Maßnahme umfasst das Lernsystem Hochschule als Ganzes mit ihren verschiedenen Akteuren und deren unterschiedlichen Interessen und Zielen. Entsprechend der Vielfalt der Hochschulangebote zu interkultureller Kommunikation und Kompetenz variieren die kontextuellen Rahmenbedingungen erheblich. Zu den kontextuellen Einflussfaktoren gehören u.a. die Organisationsform des jeweiligen Studienangebotes, die zur Verfügung stehenden Ressourcen, die Ausgestaltungsmöglichkeiten des Lehr- und Lernarrangements, die Strukturen der verantwortlichen Hochschuleinheit sowie die Beteiligung diverser Akteursgruppen (vgl. Kromrey 2000).

- *Entstehungsziele:* Eine Evaluation kann immer auch dann durchgeführt werden, wenn zum Zeitpunkt der Einschätzung und Bewertung ein konkretes Konzept und Design einer bestimmten Maßnahme oder eines Programms noch nicht vorliegt und dieses erst gestaltet bzw. entwickelt werden muss. So liefert die Evaluation und Qualitätssicherung einen Beitrag zur *Konzeption*: etwa zur *Anforderungsdefinition* (unter Berücksichtigung der Anforderungen verschiedener Akteure), zur *Aufwandskalkulation* und – wo nötig – zur *Akkreditierung* von Studiengängen. Für zu evaluierende innovative Studienprogramme und Vermittlungsmethoden formulieren Rossi und Freeman (1993) verschiedene Bedingungen. Vor allem zwei der von ihnen aufgeführten Kriterien innovativer Maßnahmen lassen sich auf den hier betrachteten Kontext interkulturell ausgerichteter Studienangebote an Hochschulen übertragen: Erstens handelt es sich meist um neu aufgesetzte bzw. zu entwickelnde Studienangebote, deren Funktionieren und Wirksamkeit zwar angenommen, keinesfalls aber umfassend erprobt

wurde; dies gilt auch, wenn es sich um die Umwandlung von Diplom- und Masterstudiengängen im Rahmen des Bologna-Prozesses handelt.² Die zu vermittelnde Qualifikation ‚interkulturelle Kompetenz‘ ist in den verschiedenen Studiengangsdokumenten (Studien- und Prüfungsordnung) zudem allenfalls sehr vage definiert (Weidemann et al. 2007). Zweitens impliziert die Vermittlung interkultureller Kompetenz häufig wesentlich stärker praxisbezogene und verhaltensorientierte Lernziele als dies im Lehr-Lernsystem Hochschule (insbesondere Universitäten) gewöhnlich umgesetzt wird.

- *Durchführungsziele:* Bezogen auf die Durchführung von Qualifizierungsmaßnahmen wird häufig von Vertretern der Hochschullehrer die übergeordnete Frage gestellt, wie das Wissen über den untersuchten Bereich, seine Strukturen und Prozesse grundsätzlich vertieft und verbreitert werden kann. Bei der Qualitätsverbesserung von Studium und Lehre wird fast immer dieser *fachdidaktische Erkenntniszweck* angesprochen: Man ist daran interessiert, empirisch fundiertes Wissen über den Erfolg der Hochschullehre bzw. des Hochschulstudiums im Allgemeinen und die in diesem Kontext vermittelten interkulturellen Kompetenzen im Besonderen zu erlangen. Ein weiteres Ziel kann die Etablierung eines *Dialogs* zwischen den verschiedenen Akteuren zur Steigerung der *Transparenz* darstellen – auch mit der Absicht, Impulse zur Weiterentwicklung der Maßnahme zu gewinnen.
- *Bewertungsziele:* Fragen nach der Effektivität und Akzeptanz von Maßnahmen richten sich aus übergeordnetem Interesse verschiedener Akteure häufig auf die Steuerung und *Kontrolle in Richtung erwünschter Effekte*. Häufiges Interesse, insbesondere von Kostenträgern ist der effiziente Einsatz von finanziellen Mitteln sowie technischen, zeitlichen und personellen Ressourcen. Hier steht als Ziel die *Legitimation des Mitteleinsatzes* im Vordergrund (vgl. Stockmann 2006). Aus politischen Erwägungen geht es manchen Akteuren auch um die *Legitimation der inhaltlichen Ausrichtung*. Gelegentlich ist die Durchsetzung einer bestimmten *Entscheidung über die Zukunft einer Maßnahme* das Ziel und dazu die Veränderung und Beeinflussung von Personen oder Personengruppen in deren Einstellung oder *Meinungsbildung zur Maßnahme* kann bestimmten Akteuren wichtig sein. Ein verbreitetes übergeordnetes Ziel insbesondere in forschungsorientierten Bereichen ist *wissenschaftliche Erkenntnis* über Vorgänge, Ursachen und Zusammenhänge, die den Effekten der untersuchten Maßnahme zugrunde liegen. Die wissenschaftliche Auseinandersetzung dient

2 Als Beispiel mag hier die Umwandlung des Master-Hauptfaches „Interkulturelle Kommunikation“ in den forschungsorientierten Masterstudiengang „Interkulturelle Kommunikation – Interkulturelle Kompetenz“ an der TU Chemnitz dienen (siehe Straub/Nothnagel 2007).

dabei oft nicht nur der Optimierung der Maßnahme, sondern auch dem übergeordneten Erkenntnisstreben (Kromrey 2000). Bei den kontextbezogenen Bewertungszielen spielen die unterschiedlichen Interessen der verschiedenen Akteure eine besondere Rolle: Darum ist hier auch die offene Frage nach jedwedem *eigenen Zielen von Evaluator/innen und deren Auftraggebern* wichtig, zumal sie leicht der Forderung nach deren Neutralität entgegen läuft. Ein anderes Ziel, das den übergeordneten Kontext betrifft, ist die Anfertigung von *Meta-Evaluationen* bzw. *Evaluationsreviews* und Meta-Analysen, in denen zusammenfassende Bewertungen von Wirkungen sowie ganzheitliche Aussagen über die Effektivität der einbezogenen Interventionsmaßnahmen auf Grundlage mehrerer Evaluationen angestrebt werden.³

Subjektbezogene Ziele

Die Evaluationsziele, die die *Interessen der einzelnen Akteure* bzw. *Teilnehmer* als ‚Subjekte‘ einer Maßnahme bedienen, führen zu einer individualisierten Betrachtung des Prozesses aus subjektiver Perspektive (z.B. der Studierenden oder Lehrenden).⁴

- *Entstehungs-Ziele:* Die *individuelle Anpassung* der Maßnahme an die Lernziele und persönlichen Bedürfnisse der Lernenden stellt einen Qualitätsanspruch an die Lehre dar. Dazu kann eine individuelle *Bedarfsabfrage* oder Erwartungskklärung – wie bei individualisierten Lehrmethoden üblicherweise vorgesehen – etwa von Seiten des Dozenten dienlich sein. Implizit wird eine solche Anpassung stets durch die Natur des Lernprozesses als aktive kognitive *Konstruktion des Lerners* stattfinden. Auch durch die *Auswahl* und individuelle Zusammenstellung zur Wahl stehender alternativer Studienangebote geschieht eine implizite Abstimmung und damit auch eine indirekte Bewertung. Die Interessen von Teilnehmern erstrecken sich nicht nur auf die Ausrichtung des inhaltlichen Angebots auf ihren Bedarf, sondern auf eine *lernerfreundliche Gestaltung von Bedingungen* wie das methodische Vorgehen, Bereitstellung von Lehrmitteln, Gestaltung von Lernorten und Rahmenbedingungen in Administration, Prüfungswesen oder Zertifizierung. All dies aus Lernersicht zu bewer-

ten, um es in der Entstehung oder Verbesserung der Maßnahme zu berücksichtigen, kann ein weiteres Ziel von Evaluation sein.

- *Durchführungs-Ziele:* Für den Lerner ist das zentrale Ziel von Bewertungen die *Steuerung* ihres individuellen Lernprozesses und -fortschritts. Dafür hat Scriven (1991) ursprünglich den Begriff der *formativen* Evaluation geprägt. Dies kann durch den Dozenten in Form mehr oder weniger formalisierter Lernzielkontrollen und Feedback, durch Einschätzungen des Lerners selbst oder durch Abfragen, die durch interaktive Medien bereitgestellt werden, erfolgen. In jedem Fall muss es zu einer Einschätzung des Lernfortschritts oder einer Bewertung der inhaltlichen Interessen kommen. Auf dieser Basis kann durch den Dozenten, durch ein adaptives Medium oder durch den Lerner selbst der Lernprozess individuell angepasst werden. Die jeweils nötige Entscheidungsgrundlage kann durch subjektive Einschätzungen oder aber auch explizite Bewertungen (Tests, Lernzielkontrollen, etc.) geschaffen werden. Teils gilt es auch die Handlungen bzw. *Handlungsorientierungen* von Teilnehmern in studienbezogenen Entscheidungssituationen (etwa bei der Wahl von Studienfächern oder Spezialisierungen) durch Rückmeldung zu steuern (z.B. Wottawa/Thierau 2003). Neben der Förderung gewünschter und der Korrektur unerwünschter Lerneffekte kann auch die Ausrichtung und Begünstigung der *Motivation* der Teilnehmer Ziel von Bewertung sein. Ein weiteres Ziel betrifft den *Transfer*: der Lerner kann Bewertungsergebnisse und Feedback zu seiner Person als handlungsleitende Erkenntnisse für die Umsetzung des Gelernten im Alltag nutzen. Besonders transferrelevant sind Erkenntnisse zum eigenen Verhalten und den eigenen Einstellungen, die im Kontext interkultureller Kompetenzen relevante Lernzielkategorien darstellen. Häufig werden für die Durchführungsziele dieselben Erhebungsinstrumente genutzt wie für die Bewertungsziele, was aber nicht immer im Sinne der genannten Ziele der Lernprozessessteuerung und Transfersteuerung ist.
- *Bewertungs-Ziele:* Die Rückmeldung über erreichte Lernziele an die einzelnen Teilnehmer dient nicht nur der (Selbst-)Steuerung des Lernprozesses, sondern ist unabhängig davon auch Grundlage der *Selbsterkenntnis* für den Lerner, u.a. zur Regulation und Stabilisierung seines Selbstbildes. Vielzitiertes Bewertungsziel ist die *Zertifizierung* der erreichten formalen Qualifikation. Insbesondere die Hochschule wird traditionell und auch mit breiter öffentlicher Akzeptanz als ‚Hoheit‘ zur Vergabe von Qualifizierungsnachweisen anerkannt. Die *Abschlüsse* qualifizieren die Studierenden für berufspraktische und wissenschaftliche Tätigkeiten. Zwischenergebnisse von Studienleistungen entscheiden zunächst aber auch über die *Berechtigung* zur Teilnahme an weiterführenden Lehrveranstaltungen. Instrumente für alle subjektbezogenen Bewertungsziele sind je geeignete Formen von *Prüfungen*.

³ Für Meta-Evaluationen und Evaluationsreviews zu interkulturellen Trainings siehe z.B. Ehnert 2004; Kealey/Protheroe 1996; Mendenhall et al. 2004; zu Meta-Analysen z.B. Deshpande/Viswesvaran 1992; Morris/Robie 2001.

⁴ Zu den unterschiedlichen Zielen einzelner Akteure und Akteursgruppen bzw. Teilnehmer sowie den damit verbundenen Herausforderungen im Zusammenhang mit der Lehre interkultureller Kommunikation und Kompetenz siehe ausführlich Weidemann/Nothnagel (i.d.B., Kap. 3.1).

Die verschiedenen Evaluationsziele können nicht unabhängig voneinander betrachtet werden. Bei der Entscheidung für oder gegen bestimmte Ziele sollten evidente oder potentielle Zielkonflikte berücksichtigt werden. Dies verdeutlicht auch, wie relevant es ist, überhaupt die Ziele eines Evaluationsvorhabens offen zu explizieren. Zu bedenken ist auch, wie realistisch eventuelle Maximalforderungen sind. Schließlich gilt es zu beachten, dass kein Evaluationsprojekt die Wirkung einer Maßnahme bzw. deren Effekte letztgültig identifizieren kann. Dies gilt – nicht zuletzt aufgrund der Komplexität des Konstrukts ‚interkulturelle Kompetenz‘ und der höchst diversen Anwendungskontexte – insbesondere für die Evaluation interkulturell ausgerichteter Studienangebote. Als hilfreich hat sich die Festlegung von Zwischenzielen erwiesen, an denen die an der Evaluation beteiligten Akteure ihr Handeln ausrichten können.

Evaluationsdesign

Erst nach Klärung von Gegenstand und Zielen kann das dazu passende Vorgehen einer Evaluation festgelegt werden. Die genannten Ziele, Kriterien und Indikatoren können in verschiedenen Evaluationsdesigns erfasst, spezifiziert und eingeschätzt werden, wobei, soweit möglich, auf dokumentierte Erfahrungen bei der Evaluation interkultureller Qualifizierungsmaßnahmen zurückgegriffen werden sollte. Entscheidungen über das Design bzw. die Untersuchungsanlage einer Evaluation dienen dazu, einen konzeptionellen Rahmen zur Planung, Durchführung und Implementierung von Evaluationsmaßnahmen zu schaffen, der eine methodische und theoretische Begründung der Evaluation liefert und meist auf praktischen Evaluationserfahrungen beruht (Beywl 2006), wobei meist unterschiedliche Designs miteinander kombiniert oder ineinander integriert werden können (vgl. Blake/Heslin 1983; Tashakkori/Teddlie 2003).

Zu solchen Designs von Evaluationen, die sich vorwiegend am „positivistischen Evaluationsparadigma“ (Ehnert 2007) orientieren, zählen u.a. experimentelle und quasi-/nicht-experimentelle Designs, Zeitreihenanalysen, Querschnitt- und Längsschnittuntersuchungen, Pretest-Posttest-Designs, Panel-Untersuchungen sowie Meta-Analysen bzw. Meta-Evaluationen (vgl. zum Überblick Mendenhall et al. 2004). In der einschlägigen Literatur wird die Gestaltung einer Evaluation in Form eines quasi-experimentellen Designs deutlich anderen Untersuchungsplänen vorgezogen (vgl. z.B. Beywl 2006; Meyer 2007). Wenn demgegenüber der „Gesprächsscharakter der Evaluation“ (Bohnsack 2006) bzw. deren „moderierende Funktion“ (Kardorff 2007) betont werden soll, z.B. in Form eines gemeinsamen Dialogs zwischen den Evaluatoren/innen und den intendierten Nutzern der Evaluationsergebnisse über ähnliche und differente Werthaltungen, Handlungsorientierungen und Bewertungen, kann mitunter auch das Verfahren einer „responsiven Evaluation“

(Stake 2004) zum Einsatz kommen. Dabei handelt es sich streng genommen um eine methodische Konkretisierung des kontextbezogenen Durchführungsziels „Dialog zwischen den Akteuren und Teilnehmern“, wenn es dabei „im Wesentlichen darum [geht; M.A./T.M.], den beteiligten Gruppen ein gegenseitiges Verständnis für unterschiedliche Erfahrungen, Perspektiven, Bewertungen und Werthaltungen zu ermöglichen“ und den/die Evaluator/in dazu zu beauftragen, „der Öffentlichkeit und vor allem den Programmbeteiligten Einblicke in die Standortgebundenheit der Perspektiven der jeweils anderen Beteiligten(-gruppen) zu eröffnen“ (Nentwig-Gesemann 2006: 164). Einen wichtigen Beitrag zu einer qualitativen, empirisch fundierten Hochschullehrevaluation vermögen insbesondere die letztgenannten Modelle in Verbindung mit rekonstruktiven Verfahren der Sozialforschung zu liefern (z.B. Bohnsack 2008).

Erhebungs- und Auswertungsinstrumente

Mit der Entscheidung für eine konkrete Evaluationsmethodik bzw. ein Evaluationsdesign ist untrennbar auch die Wahl geeigneter Erhebungs- und Auswertungsinstrumente verbunden, welche an den Evaluationszielen bzw. der zu untersuchenden Fragestellung und dem Evaluationsdesign auszurichten sind. Im Rahmen von Evaluationen werden bekannte Methoden und Verfahren der empirischen Sozialforschung eingesetzt. Dabei können quantitative Methoden, die dem induktiv-statistischen Erklärungsmodell folgen und nach Generalisierbarkeit und Objektivität, Reliabilität und Validität der gefundenen Aussagen streben, von qualitativen Methoden, die dem interpretativen, hermeneutisch-verstehenden Erklärungsmodell folgen und auf eine bereichsspezifische Vertiefung des Orientierungs- und Handlungswissens über einen Forschungsgegenstand abzielen, unterschieden werden (vgl. Kochinka/Werbik 2000). Als geeignete und für die Erforschung interkultureller Kommunikation und Kompetenz sensible Methoden und Verfahren erscheinen u.a. Befragungen und Interviews, Fallanalysen und Analysen kritischer Interaktionssituationen, Experimente und Beobachtungen. Zur Datenauswertung können z.B. Verfahren der deskriptiven, multivariaten (z.B. Faktoren- und Clusteranalyse) und Inferenzstatistik (vgl. Diekmann 2007), aber auch gesprächs- und konversationsanalytische Textanalysemethoden, qualitative Inhaltsanalysen und Textinterpretationsverfahren im Rahmen der interpretativen, rekonstruktiven Sozialforschung (vgl. Bohnsack 2008) eingesetzt werden.

Im Kontext der Evaluation interkulturell ausgerichteter Studienangebote kann auf eine Vielzahl von Testverfahren und Instrumente zur diagnostischen Erfassung interkultureller Kompetenz zurückgegriffen werden, die sich jedoch hinsichtlich ihrer theoretischen Fundierung sowie der Validität und Reliabilität erheblich unterscheiden (Deller/Albrecht 2007). Bisherige Metaevaluatio-

nen spiegeln nur ein unzureichendes Bild von der Vielfalt an einsetzbaren Methoden wider. Weitgehend vernachlässigt werden qualitative Ansätze, Verfahren und Methoden, die einen empirischen Zugang zu biografisch und kollektiv geprägten Erlebnissen, Erfahrungshintergründen und Handlungsweisen von Menschen, zu für ihre Alltagspraxis relevanten *common-sense*-Theorien und impliziten (kulturellen), praktischen Orientierungswissen eröffnen können. Das Konzept einer qualitativen Evaluationsforschung ist schließlich in Form einer empirisch gestützten Beschreibung und Analyse von Erlebnis- und Handlungsfeldern der beteiligten Akteure denkbar, wenn damit das Ziel verfolgt wird, nicht nur die institutionelle Praxis des Evaluierens selbst zu untersuchen, sondern auch die alltäglichen Praktiken, die subjektiven Theorien und Diskurse aller beteiligten Akteure in die Evaluation einzubeziehen. Dabei kommen „qualitative, prozess-, kommunikations- und beteiligungsorientierte Verfahren“ zum Einsatz (Kardorff 2007: 242), ohne dass die Ansprüche der Effizienzquantifizierung und summativen Evaluation dadurch ersetzt werden. Vielmehr werden letztere durch Einbezug der Interessenlagen und Meinungen der Betroffenen und durch eine weniger ergebnisbezogene, dafür aber ablaufbegleitende, formative Evaluation ergänzt und erweitert. Viele Autoren sprechen sich daher bei der Evaluation interkultureller Interventionsmaßnahmen dafür aus, sowohl quantitative als auch qualitative Verfahren miteinander zu kombinieren (vgl. Mendenhall et al. 2004). Nur so kann schließlich neben der Frage, ob und wie interkulturell ausgerichtete Studienangebote bestimmte Wirkungen erzielen, auch geklärt werden, warum sich bestimmte Effekte bei der Entwicklung interkultureller Kompetenz überhaupt erst entfalten, die sowohl als signifikant positive als auch negative Wirkungen messbar sind und von den an einer Maßnahme Beteiligten relevant gesetzt werden. Durch die Berücksichtigung der individuellen Perspektiven und qualitativen Bedeutung der Teilnehmer- bzw. Akteursperspektive können schließlich auch die Transfer- und Anwendungsfähigkeit von in der Hochschullehre vermittelten Aspekten interkultureller Kommunikation und Kompetenz auf das fremdkulturelle Handlungs- und Lebensumfeld besser eingeschätzt werden. Mit der Wahl der anzuwendenden Erhebungsinstrumente sowie des einzusetzenden Design wird im gleichen Moment auch eine Entscheidung über den Zeitpunkt der Evaluation, die Beteiligten sowie über die Einhaltung von vorab definierten Evaluationsstandards und Gütekriterien für die Bewertungsvorgänge getroffen.

Evaluationszeitpunkt

Üblicherweise wird bei der Bestimmung der Erhebungszeitpunkte und damit des Ablaufs von Evaluationsmaßnahmen die begriffliche Unterscheidung zwischen formativer und summativer Evaluation in Anspruch genommen.

Hense (2006) weist darauf hin, dass die originäre Bedeutung dieser Begriffe als „intendierte Evaluationszwecke“ beachtet werden sollte und es keinerlei Sinn mache, den Zweck der Evaluation einerseits und den Zeitpunkt oder Evaluationsgegenstand andererseits *per se* gleichzusetzen, was fälschlicherweise suggerieren würde, dass damit die Vielfalt möglicher Formen der Evaluation abgebildet werden könne. Einem ähnlichen Vorbehalt unterliegen auch andere Unterscheidungen, wie z.B. in eine ‚ergebnisorientierte‘ vs. ‚prozessbegleitende‘ Evaluation (Kardorff 2007).

Trotz dieser begrifflichen Unschärfe kann danach gefragt werden, wann bzw. zu welchen (Mess-)Zeitpunkten Bewertungshandlungen stattfinden sollten. Mendenhall et al. (2004) haben hinsichtlich des Evaluationszeitpunktes gezeigt, dass im Falle der Evaluation interkultureller Interventions- und Qualitätsmaßnahmen die abhängigen Variablen häufig vor und nach der Maßnahme zur Vermittlung interkultureller Kompetenz, selten auch während der Durchführung des Trainings gemessen wurden. Dieses Verständnis von Erhebungszeitpunkten wurde vielfach kritisiert, denn Evaluationsmaßnahmen können und sollten in allen Stadien eines Lern- und Entwicklungsprozesses ansetzen, um dem Anspruch zu genügen, Qualität zu sichern und Effekte zu erklären, statt nur Wirkung festzustellen. Demnach sollte eine Evaluation bereits vor Durchführung eines Programms, möglichst schon im Entstehungs- und Konzeptionsprozess eines Lernsystems, also mit der Ermittlung der Anforderungen und des Bildungsbedarfs der Zielgruppe, beginnen und auch die daran anschließende Zusammenstellung der zu vermittelnden Lerngegenstände und die Setzung der konkreten Lernziele sowie die Auswahl der anzuwendenden Vermittlungsmethoden sowie die Wahl bzw. Erstellung geeigneter Vermittlungsmedien – mithin den gesamten Prozess der Entstehung einer Maßnahme – umfassen. Eine Evaluation sollte nicht allein als abschließende Tätigkeit am Ende eines Gestaltungs- und Entwicklungsprozesses verstanden werden, sondern auch als kontinuierliche, prozessbegleitende wissenschaftliche Handlung, die den gesamten „Lebenszyklus“ des Lernsystems auf Makro-, Meso- und Mikroebene als Evaluationsgegenstand begleitet (Mayer 2001: 327ff.). Bewertung findet dann nicht nur nach Entstehung und Durchführung des Gesamtprogramms, sondern auch nach den kleineren Untereinheiten bis hin zur Einzelinteraktion statt (s.o.). Ein solches Vorgehen wird der Forderung nach einer prozessbegleitenden Evaluation gerecht. Damit können sowohl ergebnisorientierte und prozessbegleitende Aspekte, Komponenten und Bausteine einer Interventionsmaßnahme, als auch einzelne Lernelemente bzw. Lernschritte und deren Qualitäten in Abhängigkeit von vordefinierten und zu erreichenden Lernzielen evaluiert werden (siehe Blüml i.d.B. Kap. 3.3). In diesem Sinne beschränkt sich Evaluation nicht nur auf Ziele der Bewertung der durchgeführten Maßnahme zum Zwecke ihrer Veränderung oder Verbesserung für neuerliche Durchführungen, sondern kann als eine alle Be-

standteile, Bausteine und Entwicklungsschritte umfassende ganzheitliche Begleitforschung verstanden werden.

Evaluator/in bzw. Evaluatorenteams

Die Rolle der Evaluator/innen bzw. Evaluatorengruppen im Evaluationsprozess wird durch verschiedene Rahmenbedingungen beeinflusst (Stockmann 2006: 263f.): (1) ihre institutionelle Einbettung in die Evaluationsmaßnahme; (2) ihr implizites Rollenverständnis und Wissen über die Planung, Durchführung und Implementierung geeigneter Evaluationsstrategien; (3) die Ziele der Evaluation.

Je nach Ort der Durchführung und Steuerung sowie institutionellen Einbindung einer Evaluationsmaßnahme kann zwischen einer internen (durch die eigene Organisation) und externen (durch einen von der Organisation beauftragten Dienstleister) Evaluation unterschieden werden (Brandt 2007; Widmer 2000). Wird für die Evaluation ein Rollenverständnis der Evaluatoren angestrebt, welches durch Unabhängigkeit, Neutralität und Objektivität sowie eine professionelle Spezialisierung auf methodisch und wissenschaftlich fundierte Evaluationsdienstleistungen im Zusammenhang mit interkulturell ausgerichteten Weiterbildungsmaßnahmen gekennzeichnet ist, muss i.d.R. auf externe Evaluatoren mit spezifischer Expertise im Bereich ‚interkulturelle Kompetenz‘ zurückgegriffen werden. Während bei den Dozenten der zu evaluierenden Studienangeboten Qualifikationen für und Erfahrungen mit Evaluationsmaßnahmen nicht immer vorausgesetzt werden können, mangelt es hochschulinternen Evaluatoren meist an Fachkenntnis hinsichtlich des Gegenstandsbereichs ‚interkulturelle Kompetenz‘. Interne Evaluator/innen stehen zudem im Risiko, tendenziell eigenen Zielen und Interessen bzw. denen der Organisationen, der sie angehören, zu folgen.

Dennoch stellen Varianten der *Selbstevaluation* eine möglicherweise nützliche Option dar (Hense 2006). Die Selbstevaluation bildet einen Spezialfall einer internen Evaluation, bei der unmittelbar und mittelbar an einer Evaluation beteiligte bzw. von der zu evaluierenden Maßnahmen betroffene Personen und Interessensgruppen mitwirken und so als „owners of the process“ (ebd.) einen wesentlichen Einfluss auf Evaluationsziele, Fragestellungen, Design, Umsetzung und Verwendung von Evaluationsergebnissen nehmen können. Eine solche dialogische Evaluationspraxis ist jedoch begleitet von einer Rollenkonfusion der beteiligten Akteure, die zugleich Auftraggeber, Entscheider und Evaluatoren sind (ebd.). Kritiker wenden ein, dass die mit der Subjektfokussierung verbundene Subjektivität die Aussagekraft und Qualität der Ergebnisse und Schlussfolgerungen der Evaluationen gefährde (z.B. Kähler 2006). Zu den Vorteilen der Selbstevaluation gehören neben den schon genannten auch der geringere Aufwand sowie die Integration des Bewertungs-

und Qualitätssicherungsprozesses in den Lehr-Lern-Prozess selbst (Arzberger/Mayer 2004). Wenn im Rahmen einer solchen „integrierten Evaluation“ (ebd.) Evaluationsergebnisse zu Entstehung und Durchführung (einschließlich erreichter Lernziele) direkt von den Akteuren, die ein interkulturell ausgerichtetes Studienangebot konzipieren und durchführen, selbst erarbeitet und interpretiert werden, ermöglicht dieses Vorgehen die unmittelbarste und schnellste Umsetzung konkreter Verbesserungs- und Veränderungsprozesse. Zudem kann man davon ausgehen, dass die betroffenen (und an der Selbstevaluation beteiligten) Akteure bzw. Teilnehmer der untersuchten Maßnahme als ‚praktische Experten‘ die Ergebnisse zu Lernerfolgen (und Misserfolgen) im Detail am besten und mit klarer eigener Motivation zur Veränderung interpretieren können. Rückmeldungen nach Abschluss einer (Fremd-)Evaluation wirken dagegen – wenn überhaupt – nicht derart direkt auf die zu untersuchende und optimierende Maßnahme zurück. Um die Effizienzvorteile der Selbstevaluation nutzen zu können, Transparenz zu ermöglichen und Nachteile (z.B. den Einfluss vordergründiger Eigeninteressen) zu kontrollieren, sollten die Zielsetzung, wesentlichen Fragestellungen, das Design und Vorgehen und vor allem auch die weitere Ergebnissichtung und -verarbeitung jedoch in transparenter Zusammenarbeit mit einer übergeordneten bzw. unabhängigen Instanz vorgenommen werden.

Im Kontext der Hochschullehre kann eine integrierte Evaluation nicht nur die Einschätzungen, je unterschiedlichen Zielvorstellungen und Interessen der verschiedenen Akteure (Dozenten, Studierende, Tutoren, Lehrstuhlinhaber, Hochschulverwaltung, Wissenschaftsministerium etc.) berücksichtigen, sondern diese Personen und Gruppen übernehmen in angemessener Weise auch die Funktionen von (Selbst-)Evaluatoren.

Selbstevaluationen – aber auch alle anderen internen und externen Evaluationen können auf unterschiedlichen organisationalen Ebenen ansetzen:

- bei der *Auswahl von Studierenden* für die Studienangebote: Daran sind neben den Studierenden selbst u.a. die Studiengangsberater der involvierten Fachbereiche, die Universitätsverwaltung in Form der Studentensekretariate sowie externe Dienstleister, wie z.B. die Zentralstelle für die Vergabe von Studienplätzen, beteiligt;
- auf *Ebene der Konzeption und Durchführung von Studienangeboten*: Das betrifft unmittelbar die Studierenden, Dozent/innen und Studiengangsleiter/innen, außerdem die Studienberater und ehemaligen Studierenden;
- auf *Ebene der Hochschule*, z.B. vertreten durch hochschulpolitische und hochschuladministrative Gremien wie etwa Fakultätsleitung, Rektorat bzw. Präsidium und Senat;
- auf *Länder- und Bundesebene*, wie z.B. Kultusministerien, Behörden und Gremien auf Bundes- und Länderebene, die mit hochschulrechtlichen Fragen betraut sind. An einer ideellen und finanziellen Förderung von

Hochschulen in Lehre und Forschung sind ebenso politische, wirtschaftliche und wissenschaftliche Interessengruppen und Verbände interessiert, z.B. Studien- und Forschungsstiftungen, Parteien und Wirtschaftsunternehmen, die Studien- und Forschungsstipendien oder Praktikantenstellen vergeben sowie als Arbeitgeber gegenüber Absolvent/innen auftreten.

Evaluator/innen müssen zwischen den vielfältigen Interessen der unterschiedlichen Stakeholder vermitteln können und ggf. auch in der Lage sein, negativen Einstellungen gegenüber Evaluationshandlungen gegenzusteuern. Nicht immer sind alle an einem Studienangebot beteiligten Akteure und Gruppen gleichermaßen an einer Bewertung durch Evaluation interessiert; mitunter fürchten sie auch negative Konsequenzen bezüglich des zu evaluierenden Studienangebots und/oder ihrer Rolle. Als zielorientiertes und systematisches wissenschaftliches Handeln, das individuelle, institutionelle und politische Aspekte in der Hochschule berührt und Wirkungen auf diese anstrebt und ausübt, ist Evaluation zudem stets in einen höchst sensiblen Rahmen ethischer und moralischer Verantwortung gegenüber allen an der Evaluation Beteiligten und der Öffentlichkeit eingebettet (zum verantwortungsbewussten Handeln von Evaluator/innen s.u.).

Teilnehmer und Stakeholder

Evaluationen in komplexen Lehr-Lernsystemen wie im Kontext der Hochschulbildung zielen nicht nur auf zu evaluierende Gegenstände und Vorgänge (z.B. Studienangebote, Vermittlungsmethoden), sondern ggf. auch auf die an der zu evaluierenden Maßnahme beteiligten Personen. Sowohl als Befragte bzw. ‚Objekte‘ der Evaluation, wie auch als Teilnehmer der evaluierten Maßnahme sind sie mit ihren individuellen, mitunter divergierenden Perspektiven und Sichtweisen, Einstellungen und Handlungsweisen involviert. An Evaluationen sind darüber hinaus verschiedene weitere Interessengruppen bzw. ‚Stakeholder‘ direkt oder indirekt beteiligt oder von dieser betroffen. Da auch sie die Planung, Durchführung und Implementierung begleiten, behindern oder fördern können (Brandt 2007: 165f.), müssen sie bei der Evaluation ebenfalls berücksichtigt werden. Mit einem weiter gefassten Stakeholder-Begriff lässt sich neben den unmittelbaren Auftraggebern und Finanziers außerdem eine weitere große heterogene Gruppe mit mittelbarem Interesse an der Konzeption, Durchführung und Evaluation von interkulturell ausgerichteten Studienangeboten identifizieren. Hierzu zählen neben nationalen und internationalen Behörden und Einrichtungen (z.B. Behörden und Gremien auf Bundes- und EU-Ebene) auch politische, wirtschaftliche und wissenschaftliche Interessengruppen und Verbände (z.B. Parteien und Wirtschaftsunternehmen, Studien- und Forschungsstiftungen), die nicht nur unmittelbar an der finanziellen und

ideellen Förderung von Forschung und Lehre zu interkultureller Kommunikation beteiligt und an der Vergabe von Praktikumsstellen bzw. Arbeitsplätzen für Absolvent/innen mit interkultureller Kompetenz interessiert sind. Diese Stakeholder sind nicht zuletzt maßgeblich daran beteiligt, den Bedarf an interkultureller Kompetenz vorzugeben und Studieninhalte mitzubestimmen – wobei die Auffassungen von interkultureller Kompetenz sich durchaus von derjenigen der Gegenstandsexperten unterscheiden kann (Weidemann/Nothnagel i.d.B., Kap. 3.1).

Die Aufgaben und Verantwortungsbereiche der Stakeholder bedürfen im Rahmen der Evaluation einer weiteren Differenzierung (vgl. z.B. die Checkliste von Fitzpatrick/Sanders/Worthen 2004). Diese Akteure können sowohl als Einzelpersonen involviert oder als informelle (z.B. eine Praktikumsgruppe von Studierenden) bzw. formelle Gruppen (z.B. Gremien wie Fachschaftsräte, Senat, Fakultätsrat) in Erscheinung treten. Die Akteure im Handlungsfeld Hochschule bzw. die Stakeholder einer Evaluation können auf unterschiedliche Weise als Informationslieferanten bzw. ‚Datenquellen‘ bei der Beurteilung von Lehrmethoden und Studienangeboten dienen: (1) Als professionelle Experten in der Hochschule verfügen sie über ein implizites Wissen darüber, wie das Lehr-Lernsystem Hochschule zu funktionieren hat; (2) sie können ihre persönlichen Eindrücke, Meinungen und Einstellungen gegenüber einer Maßnahme sich selbst und anderen Akteuren schildern; (3) Akteure definieren meist unterschiedliche Zielerreichungskriterien; (4) bestimmte Stakeholder können auch als fachliche und wissenschaftliche Experten (z.B. Informationslieferanten, Experten in der Vermittlung interkultureller Kompetenz in der Hochschullehre) fungieren.

Je nach Evaluationsziel und Evaluationsdesign muss – neben der Zusammenstellung des Evaluatorenteams – auch eine Auswahl der Teilnehmer als Evaluationsobjekte getroffen werden, welche dazu beiträgt, den Gegenstand der Evaluation zu präzisieren und zu spezifizieren. Idealerweise werden auch die Stakeholder (bzw. einzelne Vertreter) in die Evaluation integriert. Bei der Auswahl der Evaluationsteilnehmer können verschiedene Datenerhebungstechniken eingesetzt werden, z.B. repräsentative Stichproben oder theoretische, zweckorientierte Auswahlen. So können von der Durchführung, den Ergebnissen und Verwendungszusammenhängen von Evaluationen durchaus Personen und Gruppen betroffen sein, die bewusst oder unbewusst von der Evaluation ausgeschlossen worden sind. Wenn beispielsweise Studierende über Schwierigkeiten bei der studienbezogenen Planung und Durchführung oder Anrechnung von Auslandspraktika und -semester berichten, dann sollten bei der Evaluation ebenso auch vormals vielleicht nicht im Evaluationsdesign berücksichtigte Akteure in der Hochschule einbezogen werden, wie z.B. Studiensekretariate, Prüfungsämter und akademische Auslandsämter (siehe Nothnagel i.d.B., Kap. 4.14).

Evaluationsstandards

Die Beachtung von Evaluationsstandards wird i.d.R. aus zwei miteinander verwobenen Perspektiven diskutiert: zum einen unter dem Aspekt der Integration von Qualitätsansprüchen und Gütekriterien beim Einsatz von sozialwissenschaftlichen Methoden, zum anderen aus Sicht professioneller und wissenschaftlicher Verbände, wie z.B. der Deutschen Gesellschaft für Evaluation und dem Joint Committee on Standards for Educational Evaluation, die Standards und Gütekriterien für Evaluation formuliert haben. In der Auseinandersetzung mit der Güte und Wissenschaftlichkeit empirischer Sozialforschung lassen sich verschiedene Forschungspositionen erkennen (vgl. Steinke 2007). Auf der einen Seite bemüht man sich im Hinblick auf qualitative Evaluation um die Weiterentwicklung und Ausweitung traditioneller Gütekriterien quantitativer Sozialforschung auf die qualitative Forschung (vgl. u.a. Kirk/Miller 1986). Andererseits wird versucht, die Aussagekraft der erhobenen Daten nach anderen als den traditionellen Kriterien (Objektivität, Reliabilität, Validität) zu beurteilen. Bezogen auf den Prozess der Kommunikation und Interaktion der Beteiligten an einer Evaluation kann man mit Hilfe der „kommunikativen Validierung“ (vgl. u.a. Kvale 1995) die Aussagekraft der Ergebnisse erhöhen, wenn man diese den Forschungspartnern vorlegt und gemeinsam mit ihnen Reichweite, Vollständigkeit und Breite der Analysen diskutiert. Dies wird z.B. durch externe Audits von Experten und durch die Einbeziehung unterschiedlicher Interessengruppen bzw. Stakeholders ermöglicht (vgl. Kardorff 2007). Zusätzlich erscheint auch die multimethodale und -perspektivische Strategie der Triangulation sinnvoll (vgl. Flick 2004). Dabei werden verschiedene Datenquellen (wie z.B. Selbst- und Fremdb Berichte der Akteure und reflektierte eigene Erfahrungen der Evaluatoren), Untersuchungsmethoden (wie z.B. offene Interviews, Gruppendiskussionen, moderierte Zielfindungs- und Ergebnissgespräche, Dokumentenanalysen und Verfahren der ethnografischen Feldforschung) sowie Theorien der beteiligten Beobachter und Interpreten bei der Datenerhebung und -auswertung miteinander kombiniert, um die gefundenen Ergebnisse wechselseitig zu überprüfen, zu ergänzen und zu erweitern. Als allgemeines Kriterium der Wissenschaftlichkeit qualitativer Evaluationsforschung gilt ebenso die ‚intersubjektive Nachvollziehbarkeit‘, die durch die Offenlegung forschungsleitender Interessen und Vorverständnisse, Begründungen der Methodenwahl und die ausführliche Dokumentation der Analyseergebnisse sowie der einzelnen Interpretationsschritte erzielt werden soll (vgl. z.B. Steinke 2007).

Die Beurteilung der Güte von Evaluationsstudien wurde auch in der Evaluationsforschung diskutiert. Als die zwei wesentlichen Faktoren zur Einschätzung der Validität von experimentellen und quasi-experimentellen Forschungsdesigns zur Untersuchung von sozialen Handlungszusammenhängen

benennen Cook und Campbell (1979) in ihrer „Checklist“: (1) interne Validität; (2) externe Validität, Repräsentativität oder Generalisierbarkeit. Darüber hinaus zählen Hatry und Newcomer (2004) die Glaubwürdigkeit der Ergebnisse und Schlussfolgerungen, statistische Messvalidität, Validität der statistischen Erklärungsmuster und die Reliabilität bzw. Reproduzierbarkeit der Ergebnisse als Gütekriterien auf, welche die Qualität von Evaluationen beeinflussen. Meist handelt es sich aber um *trade-off*-Beziehungen. Verbesserungen der internen Validität können die externe Validität gefährden und umgekehrt (vgl. Blake/Heslin 1983).

In der Evaluationsforschung gibt es in Bezug auf die Qualitätssicherung und die Einführung von Standards der Evaluation eine gesonderte Diskussion über die Wissenschaftlichkeit derartiger Bemühungen. Cronbachs (1983) 95 Thesen bzw. Kriterien für Programmevaluationen gelten als hilfreiche Orientierung (vgl. u.a. Joint Committee on Standards for Educational Evaluation 1999). Ausgehend von einem noch inhaltlich und formalstrukturell unbestimmten Begriff der ‚Qualität‘ zählt Widmer (2000) verschiedene weitere Beurteilungskriterien für die differenzierende Betrachtung und Beurteilung qualitativer Evaluationsforschung auf: die Evaluationsgegenstände (Programm- vs. Ergebnisevaluation), die Verortung einer Evaluation (interne vs. externe; formative vs. summative; Selbst- vs. Fremdevaluation), Methodenwahl (experimentelles vs. konstruktivistisches Paradigma), die Distanzierung des/r Evaluators/-in von der Evaluation und dem Evaluationsobjekt, der Umfang einer Evaluation (zeitliche, personelle, rechtliche und finanzielle Ressourcen), Erwartungen an die Evaluation seitens der Interessengruppen Politik, Wissenschaft, Öffentlichkeit. Im Zuge der Professionalisierung der Evaluation haben sich in den 1970er Jahren – zunächst in den USA – verschiedene Initiativen gegründet, um Kriterien zur Einschätzung der Güte und Qualität einer Evaluation zu entwickeln. In Folge dessen kam es zur Formulierung von Evaluationsstandards durch das Joint Committee on Standards for Educational Evaluation, wonach u.a. ein Nebeneinander zwischen interpretativen, rekonstruktiven und klassischen, (quasi-)experimentellen Methoden angestrebt wird. Im deutschen Sprachraum blieben diese Standards lange Zeit unbeachtet. Erst Ende der 1990er Jahre kam es zur Übersetzung dieser „Standards“ (vgl. Joint Committee on Standards for Educational Evaluation 1999).⁵ Das Regelwerk benennt vier heterogene, weitgehend miteinander verflochtene

5 Die Deutsche Gesellschaft für Evaluation, die sich aus Institutionen und Personen zur Förderung der Professionalisierung und zum Informationsaustausch rund um das Thema Evaluation zusammengeschlossen hat, übernimmt in Deutschland die Herausgabe und Entwicklung von Standards (DeGEval – Gesellschaft für Evaluation e.V. 2008). Wie die ISO-Normen haben diese Standards keine rechtsverbindliche Wirkung, es sei denn, dass diese zwischen Auftraggeber und -nehmer der Evaluation vertraglich vereinbart wurden.

Bewertungskonzepte: Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit der Evaluation. Die Anwendbarkeit dieser Evaluationsstandards hat sich in der Evaluationspraxis bisher nur wenig durchgesetzt, da (1) es kaum möglich ist, stets alle Standards einzuhalten, weil diese aufgrund divergierender Zielstellungen der einzelnen Konzepte nur schwer kombinierbar und integrierbar sind; (2) aufgrund der Weite der Definition der einzelnen Standards prinzipiell Probleme bei der Anwendbarkeit im Einzelfall und dem anschließenden Transfer in die Evaluationspraxis auftreten; (3) die Gewichtung der einzelnen Standards meist unklar bleibt; (4) die Kontrollfunktion der Standards nur eingeschränkt entfaltet werden kann, weil es sich im engeren Sinne um Maximal Kriterien handelt (Stufflebeam 2003; Widmer 2000).

Für eine verlässliche (auf einem experimentellen Design beruhende) Studie zur Wirksamkeitseinschätzung von interkulturellen Qualifizierungsmaßnahmen und der dabei eingesetzten Methoden und Verfahren empfehlen Kealey und Protheroe (1996) sieben Gütekriterien bzw. Minimal Kriterien: (1) Vergleich zwischen Experimental- und Kontrollgruppe; (2) Überprüfung von kognitivem Wissen und Verhaltenskompetenz, sowohl vor als auch nach dem Training. Über diese Minimal Kriterien hinaus sollte eine sorgfältig geplante (experimentelle) Studie außerdem folgende Bestandteile beinhalten: (3) Auswahl von Testpersonen (sowohl in Untersuchungs- als auch Kontrollgruppe) auf zufälliger Basis; (4) Einsatz von Langzeitstudien, um die spätere *on-the-job*-Leistung in verschiedenen Intervallen sowohl direkt nach dem Training als auch einige Zeit später überprüfen zu können; (5) Erweiterung und Ergänzung der Selbstausskünfte der Trainingsteilnehmer um Fremdeinschätzungen aus dem nahen Umfeld (z.B. Trainer, Supervisoren, Coaches, andere Trainingsteilnehmer und Personen des Gastlandes); (6) Klarheit und Präzision bei der Bestimmung der unabhängigen Variablen (u.a. Trainingsziele, -methoden und -dauer), so dass die gemessenen Ergebnisse den entsprechenden Variablen zugeordnet werden können; (7) Klarheit und Präzision bei der Messung der abhängigen Variablen ‚Erfolg‘ und ‚Effektivität‘ und Erweiterung durch Erfassung externer Faktoren (z.B. persönliche Anpassung oder Zufriedenheit im Gastland, positive zwischenmenschliche Beziehungen mit Angehörigen der Gastkultur und die persönlichen Leistung in Bezug auf Arbeitsanweisungen). Mendenhall et al. (2004) haben im Rahmen einer Meta-Evaluation die zwischen 1988 und 2004 veröffentlichten Studien zu interkulturellen Trainings auf die Einhaltung dieser Gütekriterien hin untersucht und kamen zu dem Ergebnis: „Overall, these results show that there is not a strong fit between the research designs in the studies reviewed in this paper and the criteria for rigor in cross-cultural evaluation research“ (ebd.: 138). Dies spiegelt das allgemeine Bild der Evaluationspraxis wider, das dadurch gekennzeichnet ist, dass die geforderten methodischen Gütekriterien empirischer Sozialforschung

sowie die Standards zur Durchführung von Evaluationen nicht immer Anwendung bzw. Berücksichtigung finden.

Perspektiven und Probleme bei der Evaluation interkulturell ausgerichteteter Studienangebote

Nicht selten stößt Evaluation an ihre Grenzen hinsichtlich der Messbarkeit, Operationalisierbarkeit und Darstellbarkeit der untersuchten Wirkungszusammenhänge und Bewertungsvorgänge. Im hier betrachteten Kontext der Evaluation von Studienangeboten zur Vermittlung interkultureller Kompetenz an Hochschulen erscheinen die Potenziale noch lange nicht ausgeschöpft. Insbesondere in folgenden Bereichen gibt es noch Klärungsbedarf bzw. Verbesserungsmöglichkeiten: (1) Identifikation, Festlegung und Operationalisierung von geeigneten Indikatoren bzw. Prädiktoren bei der Lernerfolgsermittlung; (2) Berücksichtigung von kombinatorischen methodischen Arrangements in der Evaluationsforschung und Ergänzung der vorwiegenden quantitativen Evaluationsforschung durch Aspekte und Prinzipien der qualitativen, rekonstruktiven Sozialforschung; (3) Probleme der Ergebnisdarstellung und -vermittlung; (4) Fokussierung auf Transfereffekte und Anwendbarkeit der vermittelten Schlüsselqualifikationen; (5) bereichsspezifische Evaluationsforschung im Kontext der Hochschullehre.

Die Identifizierung und Operationalisierung von geeigneten Indikatoren für interkulturelle Kompetenz zur Wirkungsmessung im Lehr-Lernsystem Hochschule werden maßgeblich durch Schwierigkeiten bei der Definition, Konzeptualisierung und empirischen Fundierung des Konstrukts ‚interkulturelle Kompetenz‘ beeinflusst (vgl. Straub 2007; Blüml i.d.B., Kap. 3.3). Die geläufigen (Teil-)Komponentenmodelle und Merkmalslisten (z.B. Bolten 2000: 68; Thomas 2003: 144; vgl. Straub 2007: 41ff.) liefern zwar reichlich Ansatzpunkte für die Präzisierung des Evaluationsgegenstandes hinsichtlich seiner empirisch erfassbaren Beurteilungskriterien. Als hilfreiche Datenquelle erscheinen diesbezüglich etablierte Testverfahren und Instrumente zur diagnostischen Ermittlung, Erfassung und Entwicklung interkultureller Kompetenz (zum Überblick siehe Deller/Albrecht 2007) sowie empirisch überprüfte (interkulturelle) Lernmodelle (zum Überblick siehe D. Weidemann 2007). Die in den diagnostischen Verfahren verwendeten bzw. aus den Lernmodellen abgeleiteten Indikatoren können jedoch nicht uneingeschränkt verwendet werden: Zum einen können nicht alle Teilkomponenten gleichermaßen gut erfasst und gleichzeitig bei einer Evaluation einbezogen werden; zum anderen stellt es häufig ein schwieriges Unterfangen dar, Störvariablen und andere Einflussfaktoren zu kontrollieren bzw. zu steuern. Darüber hinaus bleibt immer das Problem, dass es keine einheitliche Definition interkultureller Kompetenz gibt und sowohl die zu evaluierenden Maßnahmen als auch die Evaluation selbst

letztlich je spezifisch definieren müssen, was jeweils als ‚interkulturelle Kompetenz‘ verstanden werden soll. Dies schränkt nicht nur die Vergleichbarkeit drastisch ein, sondern lässt – angesichts der mangelnden theoretischen Fundierung der diagnostischen Verfahren zur Messung interkultureller Kompetenz (vgl. Deller/Albrecht 2007) – auch an der (ökologischen) Validität der Ergebnisse zweifeln.

Evaluationen zur Einschätzung und Beurteilung der Erreichung von Lernzielen bzw. ‚Learning Outcomes‘ müssen letztlich auf unterschiedlichen Ebenen der Hochschulausbildung ansetzen (vgl. Blüml i.d.B., Kap. 3.3). Regelmäßig stellt es eine große Herausforderung dar, sowohl die intendierten als auch nicht intendierte Wirkungen von Lehr- und Lernmethoden zu erfassen. Die Ergebnisse der einbezogenen Evaluationsstudien zu Einschätzung der Effektivität und der Wirkungen von interkulturellen Qualifizierungsmaßnahmen können nur eingeschränkt auf die Lehrevaluation von interkulturell ausgerichteten Studienangeboten übertragen werden, da die Angebote an Hochschulen teils grundlegend anderen Charakter besitzen als z.B. interkulturelle Trainings für Mitarbeiter in Wirtschaftsunternehmen. Dennoch stellen die in diesem Kontext entwickelten Bewertungs- und Qualitätskriterien für die Entwicklung kontextsensibler Indikatoren und Erfolgsprädiktoren eine wichtige Ressource dar. Mit der Messbarmachung sind aber nicht nur Fragen der Transformation des Evaluationsgegenstands in empirisch untersuchbare Zusammenhänge, sondern auch die Wahl entsprechender und für die Evaluationsziele geeigneter Datenerhebungs- und Auswertungsmethoden verbunden.

Wie in diesem Beitrag gezeigt wurde, lassen sich Transfereffekte von sozialen Interventionsmaßnahmen, die Komplexität und Vielschichtigkeit sowie zeitliche Verzögerungen des Eintretens von intendierten Wirkungen, möglichen und beabsichtigten Nebenwirkungen sowie gegensätzlicher, ambivalenter Ergebnisse mit den bisher bekannten und eingesetzten empirischen Methoden der Sozialforschung nicht vollständig erfassen (Daumenlang et al. 1999; Ehnert 2007). Hinsichtlich der Schlüsselqualifikation interkulturelle Kompetenz sollte sich zukünftige Forschung verstärkt mit dem Transfer der kognitiven, affektiven und verhaltensorientierten Lernresultate in die Berufs- und Anwendungspraxis beschäftigen (z.B. mit den Auswirkungen des Gelernten auf das Studierverhalten und die Arbeitsleistung im Studium, während Auslandspraktika und -semester sowie im beruflichen Tätigkeitsfeld nach dem Studium). Die Transfereffekte in die Arbeitspraxis sind jedoch nur schwer zu erfassen, wie auch effektives und erfolgreiches interkulturelles Handeln meist vielschichtig, komplex und von informellen Lernprozessen abhängig ist. Bisherige Bestrebungen zur Durchführung von Evaluationen im Kontext der Hochschullehre und insbesondere zur Vermittlung interkultureller Kompetenzen bedürfen weiterhin der Verfeinerung und Vertiefung. Ein wichtiges Ziel sollte darin bestehen, bereichsspezifische Evaluationsmodelle,

-ansätze bzw. -designs zu erarbeiten, die möglichst vielen der in diesem Beitrag dargestellten theoretischen und methodologisch-methodischen Standards gerecht werden. Ein einfaches Patentrezept wird es dabei auch im Feld der Hochschule nicht geben. Allerdings lässt sich die dort angesiedelte, in vielfältigen Formen vorgenommene Vermittlung interkultureller Kompetenz differenzierter auf den Prüfstand stellen, als es in einschlägigen Evaluationsstudien bislang geschehen ist.

Literatur

- Arnold, Rolf (1992): „Evaluierungsansätze in der betrieblichen Weiterbildung“. In: Philipp Gonon (Hg.), *Evaluation in der Berufsbildung*, Aarau: Sauerländer, S. 85-108.
- Arnold, Maik (2005): „Von der Evaluation zur Evaluationsforschung. Perspektiven der qualitativen Sozialforschung“. *SIETAR Journal* 11, S. 11-14.
- Arnold, Maik (2009): *Das religiöse Selbst in der Mission: Kulturpsychologische Analyse missionarischen Handelns deutscher Protestanten*, unveröffentlichte Dissertation, Ruhr-Universität Bochum.
- Arzberger, Heinz/Mayer, Thomas (2004): „Evaluation als integraler Bestandteil der HR-Arbeit. Integrierte Evaluation beteiligt die Akteure der Personal- und Organisationsentwicklung an der Auswertung ihres Handelns“. *Personalführung* 3, S. 32-38.
- Becker-Richter, Marion/Habel, Edna/Rinke, Bernhard/Tegethoff, Hans Georg (2002): *Evaluation von Studium und Lehre. Verfahren – Methoden – Erfahrungen*, Opladen: Leske + Budrich.
- Becker, Fred G. (2005): „Den Return on Development messen. Möglichkeiten und Grenzen der Evaluation“. *Personalführung* 4, S. 48-53.
- Bergemann, Britta/Bergemann, Niels (2003): „Ausbildung interkultureller Managementkompetenz an deutschen Hochschulen“. In: Niels Bergemann (Hg.), *Interkulturelles Management*, Berlin/Heidelberg: Springer, S. 303-335.
- Bertels, Ursula/Eylert, Sabine/Lütkes, Christiana/de Vries, Sandra (2004): *Ethnologie in der Schule. Eine Studie zur Vermittlung interkultureller Kompetenz*, Münster/New York: Waxmann.
- Beywl, Wolfgang (2006): „Evaluationsmodelle und qualitative Methoden“. In: Uwe Flick (Hg.), *Qualitative Evaluationsforschung. Konzepte – Methoden – Umsetzungen*, Reinbek bei Hamburg: Rowohlt, S. 92-116.
- Beywl, Wolfgang/Schobert, Berthold (2000): *Evaluation – Controlling – Qualitätsmanagement in der betrieblichen Weiterbildung*, Bielefeld: Bertelsmann.

- Blake, Brian F./Heslin, Richard (1983): „Evaluating cross-cultural training“. In: Dan Landis/Richard W. Brislin (Hg.), *Handbook of intercultural training*, New York: Pergamon Press, S. 203-223.
- Bohnsack, Ralf (2006): „Qualitative Evaluation und Handlungspraxis – Grundlagen dokumentarischer Evaluationsforschung“. In: Uwe Flick (Hg.), *Qualitative Evaluationsforschung. Konzepte – Methoden – Umsetzungen*, Reinbek bei Hamburg: Rowohlt, S. 135-155.
- Bohnsack, Ralf (2008): *Rekonstruktive Sozialforschung. Einführung in qualitative Methoden*, Opladen: Leske + Budrich.
- Bolten, Jürgen (2000): „Interkultureller Trainingsbedarf aus der Perspektive der Problemerkahrungen entsandter Führungskräfte“. In: Klaus Götzt (Hg.), *Interkulturelles Lernen, interkulturelles Training*, München/Mehring: Hampp, S. 57-76.
- Bortz, Jürgen/Döring, Nicola (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, Berlin: Springer.
- Brandt, Tasso (2007): „Sozialer Kontext der Evaluation“. In: Reinhard Stockmann (Hg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*, Münster/New York: Waxmann, S. 164-194.
- Chen, Guo-Ming/Starosta, William J. (1996): „Intercultural communication competence. A synthesis“. *Communication Yearbook* 19, S. 353-383.
- Cook, Thomas D./Campbell, Donald Thomas (1979): *Quasi-experimentation. Design analysis issues for field settings*, Chicago: Rand McNally.
- Cronbach, Lee J. (1983): „Ninety-five theses for reforming program evaluation“. In: George F. Madaus/Michael Scriven/Daniel L. Stufflebeam (Hg.), *Evaluation models. Viewpoints on educational and human services evaluation*, Boston/MA: Kluwer-Nijhoff, S. 405-412.
- Daumenlang, Konrad/Altstötter, Christine/Sourisseaux, Andreas (1999): „Evaluation“. In: Erwin Roth/Heinz Holling (Hg.), *Sozialwissenschaftliche Methoden*, München/Wien: Oldenbourg, S. 702-713.
- DeGEval – Deutsche Gesellschaft für Evaluation e.V. (Hg.) (2008): *Standards für Evaluation*, Köln: DeGEval.
- Deller, Jürgen/Albrecht, Anne-Grit (2007): „Interkulturelle Eignungsdiagnostik“. In: Jürgen Straub/Arne Weidemann/Doris Weidemann (Hg.), *Handbuch interkulturelle Kommunikation und Kompetenz*, Stuttgart/Metzler: Metzler, S. 741-754.
- Deshpande, Satish P./Viswesvaran, Chockalingam (1992): „Is cross-cultural training of expatriate managers effective. A meta analysis“. *International Journal of Intercultural Relations* 16 (3), S. 295-310.
- Diekmann, Andreas (2007): *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*, Reinbek bei Hamburg: Rowohlt.
- Ehnert, Ina (2004): *Die Effektivität von interkulturellen Trainings. Überblick über den aktuellen Forschungsstand*, Hamburg: Kovac.
- Ehnert, Ina (2007): „Evaluation.“ In: Jürgen Straub/Arne Weidemann/Doris Weidemann (Hg.), *Handbuch interkulturelle Kommunikation und Kompetenz*, Stuttgart/Weimar: Metzler, S. 439-450.
- Fitzpatrick, Jody L./Sanders, James R./Worthen, Blaine R. (2004): *Program evaluation. Alternative approaches and practical guidelines*, Boston: Pearson.
- Flick, Uwe (2004): *Triangulation. Eine Einführung*, Wiesbaden: VS.
- Gannon, Martin J./Poon, June M.L. (1997): „Effects of alternative instructional approaches on cross-cultural training outcomes“. *International Journal of Intercultural Relations* 21 (4), S. 429-446.
- Hager, Willi/Patry, Jean-Luc/Brezing, Hermann (2000): *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien. Ein Handbuch*, Bern/Göttingen: Huber.
- Hatry, Harry P./Newcomer, Kathryn E. (2004): „Pitfalls of evaluation“. In: Joseph S. Wholey/Harry P. Hatry/Kathryn E. Newcomer (Hg.), *Handbook of practical program evaluation*, San Francisco, CA: Jossey-Bass, S. 547-570.
- Hense, Jan (2006): *Selbstevaluation. Erfolgsfaktoren und Wirkungen eines Ansatzes zur selbstbestimmten Qualitätsentwicklung im schulischen Bereich*, Frankfurt/M. u.a.: Peter Lang.
- Jenzer, Carlo (1992): „Wem nützt Evaluation – Wozu? Mit Beispielen aus dem Erfahrungshintergrund der Lehrplan- und Gesamtschulevaluation“. In: Philipp Gonon (Hg.), *Evaluation in der Berufsbildung*, Aarau: Sauerländer, S. 13-30.
- Joint Committee on Standards for Educational Evaluation (Hg.) (1999): *Handbuch der Evaluationsstandards. Die Standards des Joint Committee on Standards for Educational Evaluation*, Opladen: Leske + Budrich.
- Kähler, Harro Dietrich (2006): „Barrieren der Selbstevaluation – und wie man sie umgehen, überwinden oder schleifen kann“. *Unsere Jugend* 58 (1), S. 3-12.
- Kardorff, Ernst von (2007): „Qualitative Evaluationsforschung“. In: Uwe Flick/Ernst von Kardorff/Ines Steinke (Hg.), *Qualitative Methoden. Ein Handbuch*, Reinbek bei Hamburg: Rowohlt, S. 238-250.
- Kealey, Daniel J./Protheroe, David R. (1996): „The effectiveness of cross-cultural training for expatriates. An assessment of the literature on the issue“. *International Journal of Intercultural Relations* 20 (2), S. 141-165.
- Kirk, Jerome/Miller, Marc L. (1986): *Reliability and validity in qualitative research*, Thousand Oaks u.a.: Sage.
- Kirkpatrick, Donald L. (1979): „Techniques for evaluating training programs“. *Training and Development Journal* 33 (6), S. 78-92.
- Kochinka, Alexander/Werbik, Hans (2000): „Logische Propädeutik und Wissenschaftstheorie“. In: Jürgen Straub/Wilhelm Kempf/Hans Werbig (Hg.),

- Psychologie. Eine Einführung. Grundlagen, Methoden, Perspektiven, München: dtv, S. 42-67.
- Kromrey, Helmut (1988): „Akzeptanz- und Begleitforschung. Methodische Ansätze, Möglichkeiten und Grenzen“. *Massacommunicatio* 3, S. 221-242.
- Kromrey, Helmut (2000): „Qualität und Evaluation im System Hochschule“. In: Reinhard Stockmann (Hg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*, Opladen: Leske + Budrich, S. 233-258.
- Kvale, Steinar (1995): „Validierung. Von der Beobachtung zu Kommunikation und Handeln“. In: Uwe Flick/Ernst von Kardorff/Heiner Keupp/Lutz von Rosenstiel/Stephan Wolff (Hg.), *Handbuch qualitative Sozialforschung*, Weinheim/Basel: PsychologieVerlagsUnion, S. 427-431.
- Landis, Dan/Bennett, Janet M./Bennett, Milton J. (Hg.) (2004): *Handbook of intercultural training*, Thousand Oaks u.a.: Sage.
- Leenen, Rainer (2007): „Interkulturelles Training. Psychologische und pädagogische Ansätze“. In: Jürgen Straub/Arne Weidemann/Doris Weidemann (Hg.), *Handbuch Interkulturelle Kommunikation und Kompetenz*, Stuttgart/Weimar: Metzler, S. 773-784.
- Mayer, Thomas (2002): *I-Learning statt E-Learning: Ein integratives und universelles Modell für Lernsysteme jenseits von Schulbank und Seminarraum, Multimedia und Internet*, Dissertation, Friedrich-Alexander Universität Erlangen-Nürnberg, http://www.opus.ub.uni-erlangen.de/opus/frontdoor.php?source_opus=13&la=de, 21.01.2010.
- McDaniel, Clyde O./McDaniel, Nettie C./McDaniel, Anita K. (1988): „Transferability of multicultural education from training to practice“. *International Journal of Intercultural Relations* 12 (1), S. 19-33.
- Mendenhall, Mark E./Stahl, Günter K./Ehnert, Ina/Oddou, Gary/Osland, Joyce S./Kühlmann, Torsten M. (2004): „Evaluation studies of cross-cultural training programmes. A review of the literature from 1988 to 2000“. In: Dan Landis/Janet M. Bennett/Milton J. Bennett (Hg.), *Handbook of intercultural training*, Thousand Oaks u.a.: Sage, S. 129-143.
- Meyer, Wolfgang (2007): „Evaluationsdesigns“. In: Reinhard Stockmann (Hg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*, Münster/New York: Waxmann, S. 143-163.
- Morris, Mark A./Robie, Chet (2001): „A meta-analysis of the effects of cross-cultural training on expatriate performance and adjustment“. *International Journal of Training and Development* 5 (2), S. 112-125.
- Nentwig-Gesemann, Iris (2006): „Dokumentarische Evaluationsforschung“. In: Uwe Flick (Hg.), *Qualitative Evaluationsforschung. Konzepte – Methoden – Umsetzungen*, Reinbek bei Hamburg: Rowohlt, S. 159-182.
- Patton, Michael Quinn (2002): *Qualitative Evaluation and Research Methods*, Newbury Park u.a.: Sage.
- Rathje, Stefanie (2006): „Interkulturelle Kompetenz. Zustand und Zukunft eines umstrittenen Konzepts“. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 11 (3), <http://zif.spz.tu-darmstadt.de/jg-11-3/beitrag/Rathje1.htm>, 21.01.2010.
- Rossi, Peter H./Freeman, Howard E. (1993): *Evaluation. A systematic approach*, Thousand Oaks u.a.: Sage.
- Scriven, Michael (1991): *Evaluation Thesaurus*, Newbury Park/CA: Sage.
- Spiel, Christiane (Hg.) (2001): *Evaluation universitärer Lehre. Zwischen Qualitätsmanagement und Selbstzweck*, Münster u.a.: Waxmann.
- Stake, Robert E. (2004): *Standards-based & responsive evaluation*, Thousand Oaks u.a.: Sage.
- Steinke, Ines (2007): „Gütekriterien qualitativer Forschung“. In: Uwe Flick/Ernst von Kardorff/Ines Steinke (Hg.), *Qualitative Forschung. Ein Handbuch*, Reinbek bei Hamburg: Rowohlt, S. 319-331.
- Stockmann, Reinhard (2006): *Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement*, Münster u.a.: Waxmann.
- Straub, Jürgen (2007): „Kompetenz“. In: Jürgen Straub/Arne Weidemann/Doris Weidemann (Hg.), *Handbuch interkulturelle Kommunikation und Kompetenz*, Stuttgart/Weimar: Metzler, S. 35-46.
- Straub, Jürgen/Nothnagel, Steffi (2007): „Über das Studium interkultureller Kommunikation und Kompetenz. Akademische Anforderungen und curriculare Ausbildung in einem interdisziplinären Master-Studiengang der Universität Chemnitz“. In: Matthias Otten/Alexander Scheitza/Andrea Cnyrim (Hg.), *Interkulturelle Kompetenz im Wandel. Bd. 2: Ausbildung, Training und Beratung*, Frankfurt/M./London: IKO Verlag, S. 215-244.
- Stufflebeam, Daniel L. (2003): „Professional standards and principles of evaluations“. In: Thomas Kellaghan/Daniel L. Stufflebeam (Hg.), *International handbook of educational evaluation. Part 1: Perspectives*, Dordrecht: Kluwer Academic Publishers, S. 279-302.
- Tashakkori, Abbas/Teddlie, Charles (2003): „Major issues and controversies in the use of mixed methods in the social and behavioral sciences“. In: Abbas Tashakkori/Charles Teddlie (Hg.), *Handbook of mixed methods in social & behavioral research*, Thousand Oaks/CA: Sage, S. 3-50.
- Thierau-Brunner, Heike/Stangel-Meseke, Martina/Wottawa, Heinrich (1999): „Methoden der Evaluation von Personalentwicklungsmaßnahmen“. In: Karlheinz Sonntag (Hg.), *Personalentwicklung in Organisationen. Psychologische Grundlagen, Methoden und Strategien*, Göttingen u.a.: Hogrefe, S. 266-286.
- Thomas, Alexander (2003): „Interkulturelle Kompetenz. Grundlagen, Probleme und Konzepte“. 2. Aufl. *Erwägen – Wissen – Ethik* 14 (1), Stuttgart: Lucius & Lucius, S. 137-150.

- Weidemann, Doris (2007): „Akkulturation und interkulturelles Lernen“. In: Jürgen Straub/Arne Weidemann/Doris Weidemann (Hg.), Handbuch interkulturelle Kommunikation und Kompetenz, Stuttgart/Weimar: Metzler, S. 488-498.
- Weidemann, Doris/Weidemann, Arne/Straub, Jürgen (2007): „Interkulturell ausgerichtete Studiengänge“. In: Jürgen Straub/Arne Weidemann/Doris Weidemann (Hg.), Handbuch interkulturelle Kommunikation und Kompetenz, Stuttgart/Weimar: Metzler, S. 815-824.
- Widmer, Thomas (2000): Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In: Reinhard Stockmann (Hg.), Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Opladen: Leske+Budrich, S. 77-102.
- Will, Hermann/Winteler, Adolf/Krapp, Andreas (1987): „Von der Erfolgskontrolle zur Evaluation“. In: dies. (Hg.), Evaluation in der betrieblichen Aus- und Weiterbildung, Heidelberg: Sauer, S. 11-42.
- Wissenschaftsrat (2008): Empfehlungen zur Qualität von Lehre und Studium, vom 07.07.2008, Berlin, http://www.wissenschaftsrat.de/presse/hginfo_1408.pdf, 23.12.2009.
- Wottawa, Heinrich/Thierau, Heike (2003): Lehrbuch Evaluation, Bern/Göttingen: Huber.

Autorinnen und Autoren

Arnold, Maik, Dipl.-Kfm., ist wissenschaftlicher Mitarbeiter der Professur für Interkulturelle Kommunikation an der Technischen Universität Chemnitz. Forschungsschwerpunkte: Interkulturelle Kommunikation und Kompetenz; (Religiöse) Identitätskonzepte; missionarisches Handeln und religiöses Erfahren im Schnittfeld zwischen Religions-, Sozial und Kulturwissenschaft; Evaluation interkultureller Trainings; Methodologie und Methodik interpretativer, qualitativer Sozial- und Biografieforschung. Kontakt: maik.arnold@phil.tu-chemnitz.de.

Bauer, Gerd Ulrich, Dr. phil., Akademischer Rat im Fachgebiet Interkulturelle Germanistik (Deutsch als Fremdsprache) der Universität Bayreuth. Forschungsschwerpunkte: Auswärtige Kultur- und Bildungspolitik, visuelle Anthropologie und ethnografischer Film, Kulturbegegnung und Austauschforschung, Kulturwissenschaften, kulturwissenschaftliche Fremdeitsforschung (Xenologie). Kontakt: gerd.ulrich.bauer@uni-bayreuth.de.

Bechtel, Mark, Dr. phil., ist wissenschaftlicher Mitarbeiter der Professur für Didaktik der romanischen Sprachen der Universität Bremen. Zurzeit vertritt er die Professur. Forschungsschwerpunkte: Interkulturelles Lernen beim Lehren und Lernen fremder Sprachen, Tandem, Aufgaben- und Kompetenzorientierung im Fremdsprachenunterricht. Kontakt: bechtel@uni-bremen.de.

Blüml, Frances, M.A., ist Mitarbeiterin im Bereich Quality Management & Program Delivery der Wirtschaftsuniversität Wien, Tätigkeitsschwerpunkte: Lehr-/Lernunterstützung und -entwicklung, Feedback und Evaluierungen, Auszeichnungen in der Lehre, Programmentwicklung. Ehem. wissenschaftliche Mitarbeiterin der Professur für Interkulturelle Kommunikation an der Technischen Universität Chemnitz. Kontakt: frances.blueml@wu-wien.ac.at.